



***Facultad
de
Ciencias***

**ESTADÍSTICA DE EVENTOS EXTREMOS EN
SISTEMAS COMPLEJOS.**
(Statistics of extreme events in complex
systems)

Trabajo de Fin de Grado
para acceder al

GRADO EN FÍSICA

Autor: Cecilia Crespo Vega

Director: Juan Manuel López Martín

Octubre - 2018.

Resumen

En una serie temporal, generada a partir de un sistema natural, económico o social, donde los valores oscilan entorno un valor medio, se observan ciertos valores que exceden significativamente este nivel típico que caracteriza a la serie. Estas excedencias se conocen como eventos extremos y están relacionados con sucesos raros, es decir, sucesos cuya probabilidad de ocurrencia es muy pequeña pero sin embargo causan un gran impacto, como es el caso de terremotos, inundaciones ó crisis financieras. Consecuentemente, la necesidad de predecir este tipo de sucesos es de vital importancia en áreas como la meteorología, de cara a minimizar los daños que inevitablemente se pueden llegar a producir ó la evaluación de riesgos por parte de las compañías aseguradoras.

Con esta motivación, se ha llegado a desarrollar toda una teoría matemática destinada al estudio exclusivo de tales eventos, la cual exige separar los eventos típicos, de los verdaderos extremos dentro de una misma serie temporal. De esta manera, se podrá aplicar un modelo estadístico, conocido como la distribución de valores extremos generalizada, que agrupa a tres tipos de distribuciones: Gumbel, Weibull y Fréchet, y que ofrecerá una descripción universal de cómo se distribuyen los máximos en función de un parámetro de forma relacionado con las colas de las distribuciones de probabilidad.

El hecho de que los extremos posean propiedades universales potencia notablemente su predictibilidad, ya que da la posibilidad de adaptar un mismo modelo teórico a múltiples aplicaciones, para facilitar la inferencia estadística de las observaciones y así, la interpretación de los resultados. Por tanto, independientemente de que tipo de distribución sigan las series que tengamos inicialmente, podremos hacer una estimación de la magnitud de un evento extremo determinado o lo que es lo mismo, asignar una probabilidad de ocurrencia a dicho evento

Palabras clave: Series temporales, Eventos extremos, Teoría de valores extremos, Método de bloques, Gumbel, Fréchet, Weibull, Sistema dinámico, Mapa logístico, proceso Ornstein-Uhlenbeck, Correlación.

Overview

In a time series, generated from a natural, economic or social system, where the values oscillate around an average value, certain values are observed that significantly exceed this typical level that characterizes the series. These exceedances are known as extreme events and are related to rare events, that is, events whose probability of occurrence is very small but nevertheless cause a great impact, as is the case of earthquakes, floods or financial crashes. Consequently, the need to predict this type of event is of vital importance in areas such as meteorology, in order to minimize the damage that can inevitably occur or risk assessment by insurance companies.

With this motivation, a whole mathematical theory has been developed for the exclusive study of such events, which requires separating the typical events and the true extremes within the same time series. In this way, a statistical model, known as the generalized distribution of extreme values, grouped into three types of distributions: Gumbel, Weibull and Fréchet, can be applied, and it will offer a universal description of how the maximums are distributed according to a shape parameter related to the tails of the probability distributions.

The fact that the extremes possess universal properties greatly enhances their predictability, since it gives the possibility of adapting a single theoretical model to multiple applications, to facilitate the statistical inference of the observations and thus, the interpretation of the results. Therefore, independently what type of distribution follow the series that we have initially, we can estimate the magnitude of a certain extreme event or what is the same, assign a probability of occurrence of that event.

Key Words: Time series, Extreme events, Block Maxima Approach, Extreme value theory, Weibull, Gumbel, Fréchet, Dynamical system, Lyapunov exponent, Logistic map, Ornstein-Uhlenbeck process, Correlation.

Contents

1	Introducción	2
1.1	Series temporales	2
1.2	Eventos extremos	3
1.2.1	Teoría clásica de valores extremos	4
2	Análisis de extremos en series temporales reales	8
2.1	Block Maxima approach	8
2.1.1	Estimación de los parámetros GEV	8
2.2	Peak-over-threshold	9
3	Distribución tipo I: Gumbel	10
4	Distribución tipo II: Fréchet	13
5	Distribución tipo III: Weibull	14
6	Efectos de la correlación temporal	15
6.1	Series temporales con caos determinista	17
6.1.1	El mapa logístico	17
6.1.2	Exponentes de Lyapunov	19
6.1.3	Correlación de corto alcance	20
6.1.4	Correlación de largo alcance	21
6.2	Series temporales estocásticas	24
7	Conclusiones	28
8	Bibliografía	29
9	Apéndice	30

1 Introducción

1.1 Series temporales

El análisis de las series temporales se remonta a los estudios de las series astronómicas y climáticas que trajeron consigo, durante los primeros años del siglo XX, la formalización de la teoría de procesos estocásticos estacionarios desarrollada por los matemáticos Kolmogorov, Wiener y Cramer.

En la actualidad, la necesidad e importancia del análisis de las series temporales está presente en diferentes líneas de trabajo que abarcan varios campos de conocimiento. Todos ellos tienen en común la búsqueda de modelos que permitan la predicción del comportamiento a largo y corto plazo de diversas variables físicas, financieras, económicas y sociales para así estimar con éxito los eventos futuros de la serie en cuestión.

Para intentar resolver este problema, se hace uso de numerosas metodologías que no dejan de ser herencia de investigaciones de matemáticos, ingenieros, economistas y físicos de antaño y que van desde modelos estadísticos tradicionales, como los modelos con medias móviles, los modelos autorregresivos y los modelos tipo ARMA (combinación de los dos anteriores), hasta el uso de métodos no lineales, como los modelos de redes neuronales artificiales. Además, existen otras técnicas desarrolladas desde el punto de vista de los sistemas complejos, que serán las que usaremos en este trabajo.

Un sistema dinámico es un sistema que evoluciona a medida que avanza el tiempo y que está construido sobre un espacio de fases donde cada punto describe un cierto estado del sistema. Si el sistema evoluciona en tiempo continuo, se dice que es un sistema dinámico continuo y vendrá representado por ecuaciones diferenciales (ordinarias ó en derivadas parciales). De lo contrario, si la variable del sistema solo evoluciona en determinados instantes de tiempo, se hablará de sistemas dinámicos discretos, que vendrán definidos por mapas o iteraciones, como es el caso de la ecuación logística.

Se denomina serie temporal a una sucesión de observaciones de una variable aleatoria ordenadas secuencialmente en el tiempo:

$$X = X_0, X_1, X_2, \dots, X_n \quad (1)$$

Las series temporales pueden ser aditivas, multiplicativas ó mixtas. Consideraremos el caso más simplista, donde la serie temporal vendrá dada por un proceso aditivo, formada por una componente determinista y una aleatoria:

$$X(t) = x(t) + \zeta(t) \quad (2)$$

Sin embargo, es de interés conocer las diferencias visuales que existen entre estos dos tipos de procesos, aditivos vs multiplicativos, para ello veremos un ejemplo sencillo:

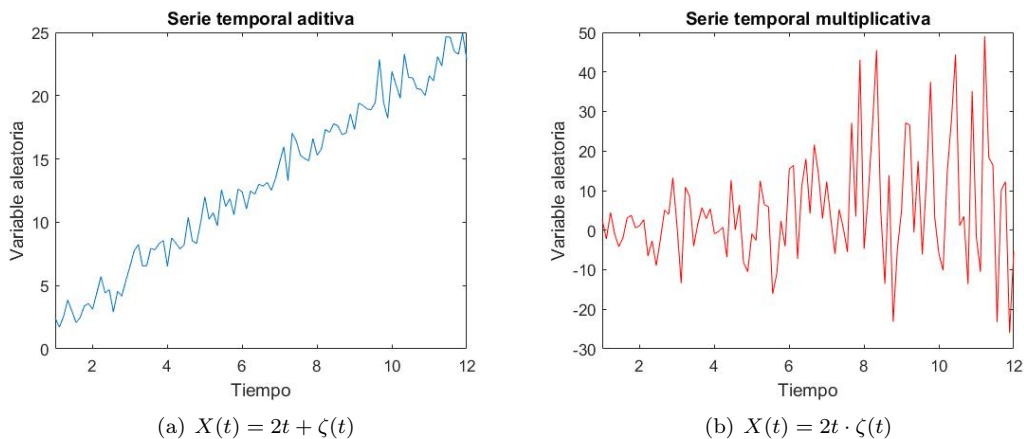


Figure 1: Ejemplo comparativo de una serie temporal aditiva y otra multiplicativa. La componente aleatoria ha sido simulada considerándola ruido blanco, es decir, un proceso estocástico gaussiano sin correlación.

En el caso del proceso aditivo, se aprecia que aunque los valores tiendan a aumentar conforme el paso del tiempo, la magnitud de los eventos dentro de un mismo periodo es aproximadamente

la misma, al contrario que en el caso del proceso multiplicativo, en donde la amplitud de las fluctuaciones aumenta a medida que transcurre el tiempo.

Otra característica importante de las series temporales es la estacionaridad. En el sentido matemático, diremos que una serie temporal compuesta por variables aleatorias $X_{i_1}, X_{i_2}, X_{i_3} \dots X_{i_k}$ es estacionaria si la distribución de probabilidad conjunta de todas las variables es la misma que la distribución de probabilidad de las variables aleatorias cuando se encuentran desplazadas un cierto periodo de tiempo t , $X_{i_1+t}, X_{i_2+t}, X_{i_3+t} \dots X_{i_k+t}$ para cualquier $k, t \in \mathbb{N}$. En el sentido físico, diremos que la serie será estacionaria si su media, varianza y por tanto, covarianza, son independientes del tiempo, es decir, la serie es estable.

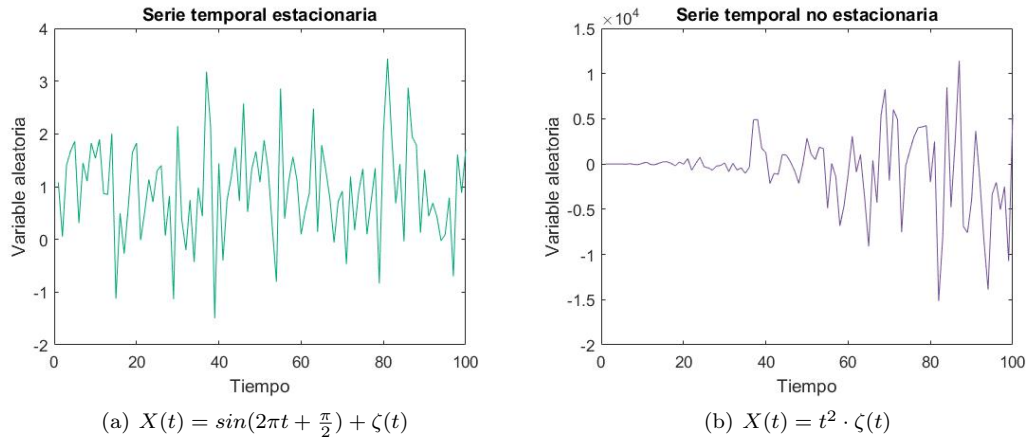


Figure 2: Ejemplos de una serie estacionaria y no estacionaria. Los procesos multiplicativos son no estacionarios ya que la varianza aumenta a medida que pasa el tiempo. El proceso aditivo mostrado anteriormente en Figure 1 también es no estacionario ya que su valor medio crece con el tiempo.

En este trabajo, todas las series que se tratan son estacionarias y en general, el hecho de saber que una serie es estacionaria resulta interesante porque poseen un mayor poder predictivo que las series no estacionarias, ya que sabiendo su comportamiento en el pasado podremos esperar con una alta probabilidad un comportamiento muy similar en el futuro.

Dado el potencial que presentan las series estacionarias existen métodos para transformar las series no estacionarias en estacionarias, este proceso se conoce como homogeneización de series temporales y se recurre a distintos métodos en función de si la serie es no estacionaria en media o en varianza. Para el primer caso, en ocasiones, se realiza una diferenciación entre cada valor y el anterior. En el segundo caso se aplican transformaciones como la logarítmica o la raíz cuadrada a todos los valores que componen la serie. Si la serie no es estacionaria ni en media ni en varianza, entonces se requiere utilizar métodos de cointegración.

1.2 Eventos extremos

Los eventos extremos son sucesos cuya probabilidad de ocurrencia es muy pequeña por lo que se sitúan en las colas de las distribuciones de probabilidad. Sin embargo, están asociados a situaciones que causan un gran impacto ya sea económico, medioambiental o social, que generalmente es negativo, como es el caso de crisis financieras, catástrofes naturales ó ataques cardiacos.

A pesar de que la probabilidad de un evento extremo decrece con su magnitud, la predicción de este tipo de sucesos en la actualidad suponen un desafío, ya que es necesario establecer precursores para determinados eventos, mientras que para aquellos que involucren escalas de tiempo mayores, es interesante ver cómo se ven afectados por los factores externos, como por ejemplo los terremotos o las inundaciones con el cambio climático.

Por tanto, el estudio de los extremos resulta imprescindible en muchas áreas de trabajo, especialmente en las geociencias y en ingeniería, para poder estimar sus tiempos de retorno y así, tratar de minimizar y medir los riesgos ante escenarios que tienen consecuencias devastadoras pero su ocurrencia, en ocasiones, resulta ser inevitable.

1.2.1 Teoría clásica de valores extremos¹

A priori, parece imposible ser capaz de predecir la probabilidad de ocurrencia de los eventos extremos y a su vez dejar abierta la posibilidad de que se traten de eventos más grandes que los ya observados con anterioridad. Para ello, es necesario separar los eventos típicos de los eventos extremos y construir un modelo estadístico que solamente se encargue de describir el comportamiento de los verdaderos extremos. La construcción matemática de la teoría clásica de los valores extremos surge de la generalización de los extremos considerándolos como eventos muy grandes y a su vez como eventos raros.

El principal resultado de la teoría clásica de extremos es el teorema de Fisher-Tippett-Gnedenko que nace como tal en 1943, y que nos dice que el máximo de una serie de variables aleatorias e independientes convergerá a una de las tres clases de distribuciones acumuladas existentes, que serán Gumbel, Fréchet ó Weibull.

Sean X_1, X_2, \dots, X_n variables aleatorias independientes e igualmente espaciadas, donde el máximo vendrá descrito como $M_n = \max \{X_1, X_2, \dots, X_n\}$. La función de distribución del máximo vendrá dada por la convolución n-ésima de la distribución:

$$\begin{aligned} P(M_n < x) &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \cdot P(X_2 \leq x), \dots, P(X_n \leq x) \\ &= F(x)^n \end{aligned}$$

Una distribución no degenerada con función de distribución $G(x)$, se dice que es una distribución de valores extremos si existen constantes $a_n \geq 0$, $b_n \in \mathbb{R}$ y además, la función de distribución es tal que:

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x)$$

Donde G es continua para cada punto x . En este caso diremos que F se encuentra en el dominio de atracción de G .

Si n es muy grande, la convergencia de probabilidades nos indicará la existencia de una distribución $G(x)$ que se ajusta a M_n .

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x)$$

Del desarrollo asintótico anterior, se llega a las tres clases para el límite del máximo y que se corresponden con la distribución de Gumbel, Fréchet y Weibull respectivamente.:

$$\begin{aligned} I : G(x) &= \exp(-\exp(-x)) & -\infty < x < \infty \\ II : G(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x < 0, \alpha < 0. \end{cases} \\ III : G(x) &= \begin{cases} \exp(-(-x)^\alpha) & x < 0, \alpha < 0 \\ 1 & x \geq 0. \end{cases} \end{aligned}$$

Las condiciones de Von Mises son condiciones necesarias y suficientes que permiten determinar el dominio de atracción G de una función de distribución F , lo que nos permitirá conocer a priori a cual de las tres distribuciones límite se ajustarán los extremos de la distribución dada. Para ello se tienen tres condiciones, una para cada distribución asintótica, de forma que en función de qué condición se satisfaga, la función F pertenecerá a un dominio de atracción u otro:

Tipo I Gumbel: Sea x_f el punto derecho final de la distribución, si $x_f = \infty$ y existe una función estrictamente positiva $g: \mathbb{R} \rightarrow \mathbb{R}$ tal que para todo $x \in \mathbb{R}$, entonces F pertenecerá al dominio de atracción de la distribución de Gumbel:

$$\lim_{h \rightarrow x_f} \frac{1 - F(h + xg(h))}{1 - F(h)} = e^{-x}$$

Además, si se tiene que $\int_0^\infty 1 - F(x)dx < \infty$ este límite se cumple y la función g puede calcularse según:

$$g(h) = \frac{\int_h^{x_f} 1 - F(x)dx}{1 - F(h)}$$

¹Todos los teoremas que se exponen en esta sección han sido consultados en la referencia [1] que aparece en la bibliografía.

Las distribuciones que pertenecerán al dominio de atracción de la distribución de Gumbel serán aquellas que tengan un soporte infinito y cuyas distribuciones decaigan más rápidamente cuando $x \rightarrow \infty$ que la distribución exponencial, por tanto, podremos esperar que para las distribuciones como la distribución gaussiana o la distribución de Poisson, los máximos se ajusten a una distribución de este tipo.

Tipo II Fréchet: F pertenecerá al dominio de atracción de la distribución de Fréchet si $x_f = \infty$ y existe $\alpha > 0$ tal que $\forall x > 0$

$$\lim_{h \rightarrow x_f} \frac{1 - F(hx)}{1 - F(h)} = x^{-\alpha}$$

Los máximos convergerán a una distribución de Fréchet cuando la distribución original tenga una cola que decrezca como una función potencial y con soporte acotado a la derecha, como es el caso de la distribución de Pareto o la distribución de Cauchy.

Tipo III Weibull: Diremos que F pertenece al dominio de atracción de la distribución de Weibull si $x_f < \infty$ y existe $\alpha > 0$ tal que $\forall x > 0$:

$$\lim_{h \rightarrow 0} \frac{1 - F(x_f - xh)}{1 - F(x_f - h)} = x^\alpha$$

Para distribuciones caracterizadas por tener un soporte acotado, como en el caso de la distribución uniforme o de la distribución beta, esperaremos que su distribución de valores extremos generalizada se corresponda con una distribución de Weibull.

Estos tres tipos de distribuciones pueden combinarse para dar lugar a un modelo unificado propuesto por Von Mises (1954) y Jenkinson (1955), denominado **distribución de valores extremos generalizada**, dependiente de tres parámetros: el parámetro de escala σ , el parámetro de localización μ y el más importante, el parámetro de forma ξ :

$$G(x; \xi, \sigma, \mu) = \exp\left\{-\left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (3)$$

donde

$$\begin{aligned} 1 + \xi \left(\frac{x - \mu}{\sigma}\right) &> 0 \\ \mu, \xi &\in \mathbb{R} \\ \sigma &> 0. \end{aligned}$$

De esta manera cuando $\xi > 0$ tendremos la distribución de tipo II ó de Fréchet, mientras que cuando $\xi < 0$ tendremos la distribución de tipo III ó de Weibull. En el caso en el que $\xi = 0$ tendremos la distribución de tipo I ó de Gumbel, que se interpreta como el límite cuando $\xi \rightarrow 0$, por lo que quedará reducida a la siguiente forma:

$$G(x; 0, \sigma, \mu) = \exp\left\{-\exp\left(\frac{x - \mu}{\sigma}\right)\right\} \quad (4)$$

Como la función de densidad de probabilidad es la derivada de la distribución de probabilidad acumulada, obtenemos **la densidad de probabilidad GEV** para los dos casos en el que $\xi \neq 0$ y para $\xi = 0$, respectivamente:

$$g(x; \xi, \sigma, \mu) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \exp\left\{-\left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \quad (5)$$

$$g(x; 0, \sigma, \mu) = \frac{1}{\sigma} \exp\left[-\left(\frac{x - \mu}{\sigma}\right) - \exp\left(-\frac{x - \mu}{\sigma}\right)\right] \quad (6)$$

El parámetro de localización μ nos da una idea de cómo de grandes son los extremos, mientras que el parámetro de escala nos indica su variabilidad. El signo del parámetro de forma ξ determina el comportamiento de las colas de las distribuciones y nos indica las propiedades cualitativas de los extremos: si el parámetro de forma es positivo, los extremos son ilimitados y tendremos colas pesadas, si el signo es negativo, los extremos tendrán un límite superior y tendremos colas muy

ligeras y poco pobladas de eventos. En el caso de la distribución de Gumbel obtendremos colas con un peso medio.

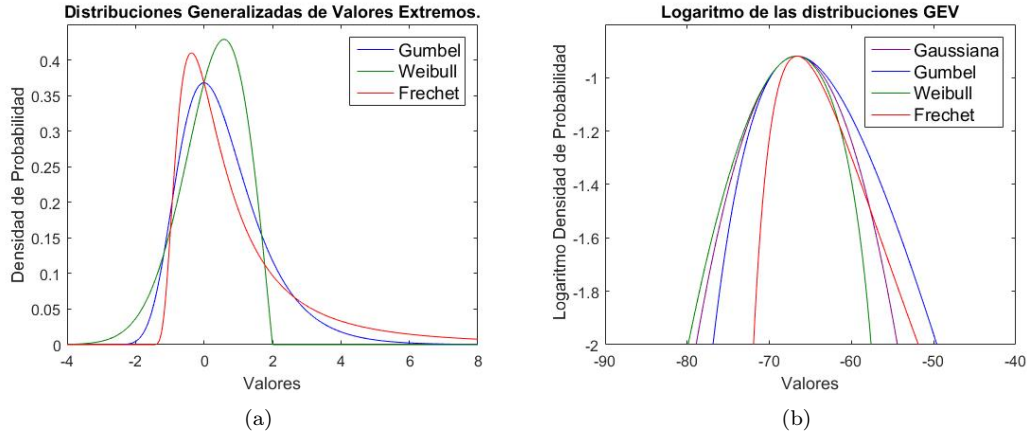


Figure 3: En ambas figuras, se ha utilizado $\mu = 0$ y $\sigma = 1$. Para la distribución de Fréchet se ha tomado $\xi = \frac{1}{2}$ y para la distribución de Weibull $\xi = -\frac{1}{2}$. En la figura (a) vemos como en función del signo del parámetro de forma, la distribución cambia, encontrando las colas más pesadas para el caso de Fréchet. Por otro lado, la figura (b) muestra en comparación con una distribución normal estándar, cómo difieren las colas de las tres distribuciones.

Como curiosidad, mencionar que existe una gran similitud entre el teorema central del límite y la teoría asintótica de los extremos de Fisher, Tippet y Gnedenko, ya que ambos van encaminados a entender el comportamiento límite de una muestra. En concreto, el teorema central del límite establece que la suma de un número dado de variables aleatorias independientes con varianza no nula tenderá a una distribución gaussiana, mientras que el teorema fundamental de la teoría clásica de extremos nos indica la familia de distribuciones a las que convergerán los extremos independientemente de como se encuentren distribuidos inicialmente.

La distribución de Gumbel fue la primera distribución propuesta para tratar de modelar los extremos y ha sido utilizada para modelar temperaturas muy altas y para predecir magnitudes de terremotos.

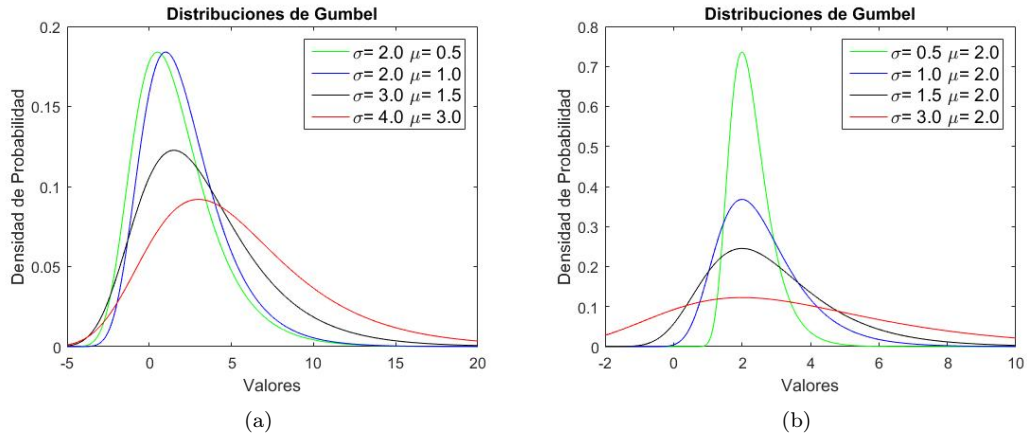


Figure 4: En la figura (a) observamos las distintas apariencias que puede presentar la distribución de Gumbel variando tanto el parámetro de localización como el parámetro de escala, mientras que en la figura (b) modificamos solamente el valor del parámetro de escala, observando que conforme mayor sea más ancha será la distribución.

La distribución de Fréchet se ha utilizado para predecir el comportamiento de flujos de protones solares y para estimar las probabilidades de eventos extremos en el DAX alemán.

Para observar los distintos aspectos que puede tomar esta distribución, vamos a modificar sus parámetros. Primero compararemos el efecto del factor de escala σ y después modificaremos solamente el parámetro de forma ξ :

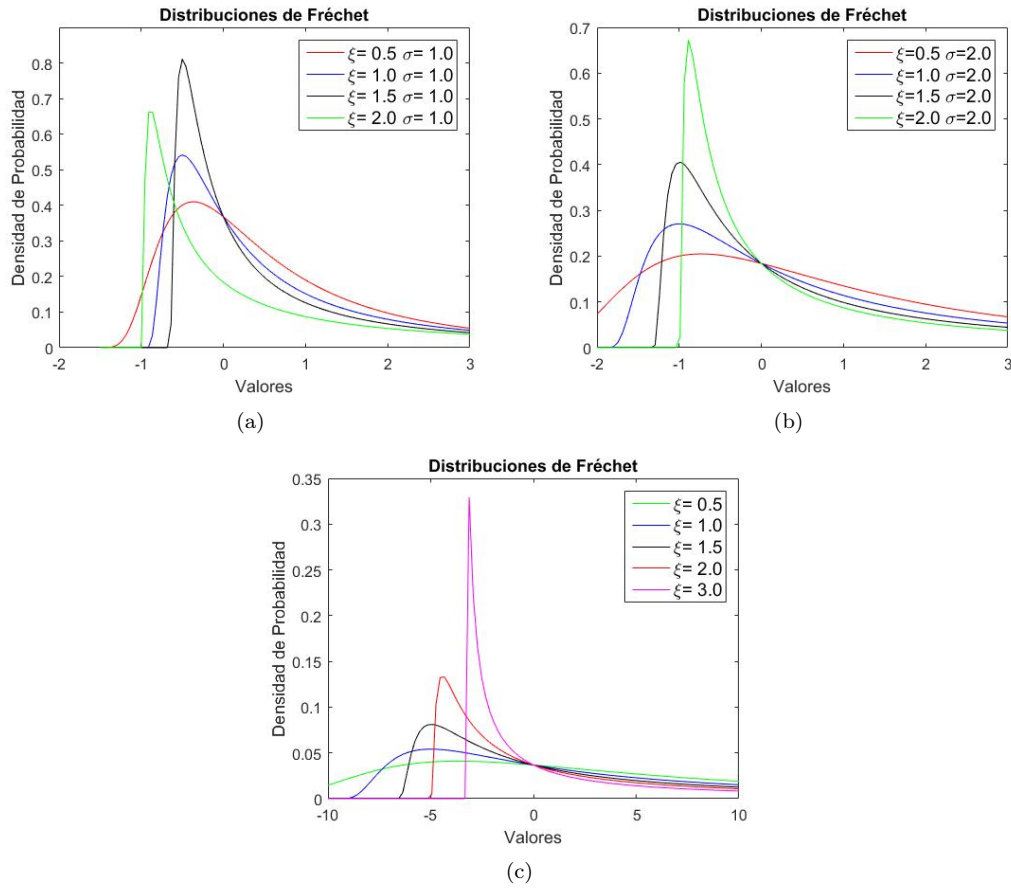


Figure 5: En las figuras (a) y (b), fijado el parámetro de localización, se puede apreciar claramente la función que desempeña el factor de escala para los mismos parámetros de forma. Mientras que en la figura (c) se puede ver como cuanto mayor es el parámetro de forma, más pesada es la cola.

La distribución de Weibull es una distribución muy versátil que puede ser utilizada para modelar distintos tipos de datos y suele aparecer con frecuencia en el estudio de control de calidades, es por ello que ha sido utilizada, por ejemplo, para analizar las fallas debidas a la corrosión de cables con revestimiento de plomo en el Centro Espacial Kennedy. Además, también ha sido utilizada para predecir la probabilidad de ocurrencia de olas gigantes en Japón y tormentas geomagnéticas.

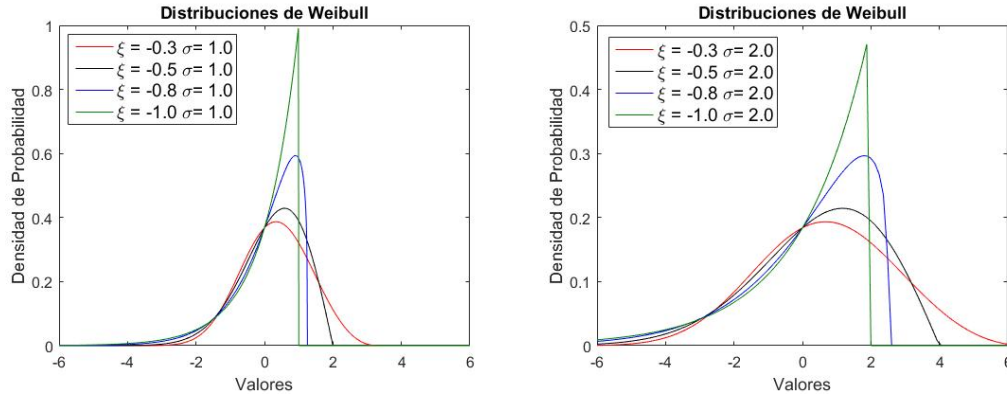


Figure 6: En la dos figuras el parámetro de localización es el mismo, y se puede apreciar cómo influyen los parámetros de escala y forma. Mientras que el primero provoca un ensanchamiento de la distribución conforme aumenta, el segundo hace que cada vez presente un aspecto más picudo.

Es importante comentar que existen dos distribuciones de Weibull. Es por ello que en ocasiones nos podemos encontrar distribuciones de este tipo con parámetros de forma positivos, cuyo uso está destinado a modelar mínimos y no máximos, como por ejemplo en casos de evaluación de mecanismos de fallas, donde el número de fallos es proporcional a una potencia del tiempo. De esta manera en función de si el parámetro de forma es mayor o menor que la unidad se tiene que el número de fallos crece o decrece, respectivamente, mientras que si toma el valor uno, las fallas producidas son constantes en el tiempo. Sin embargo, en el contexto de la teoría de valores extremos el parámetro de forma siempre será negativo y a veces podremos encontrar en la literatura, que se refieren a ella como *Weibull revertida*. Otra diferencia destacable, es que la distribución de Weibull para mínimos depende del parámetro de forma y el parámetro de escala, mientras que la distribución de tipo III de los valores extremos depende de un parámetro adicional que es el parámetro de localización.

2 Análisis de extremos en series temporales reales

Para llevar a cabo el análisis de los extremos en este TFG recurrimos a dos procedimientos: el método de las excedencias sobre un umbral (Peak-over-threshold) y el método de los bloques (Block maxima approach). Cada uno de estos dos métodos produce una serie de eventos extremos diferente con propiedades estadísticas diferentes ya que la naturaleza de los eventos extremos en ambos métodos no es la misma como se verá a continuación.

2.1 Block Maxima approach

Dada una serie temporal de longitud $s = NK$, los datos pueden agruparse en K bloques de igual tamaño, cada uno de longitud N . De cada uno de estos bloques se extrae el máximo para finalmente ajustar la distribución de valores extremos (GEV) al conjunto de los máximos asociados a cada uno de los K bloques. A pesar de ser un método sencillo, posee el inconveniente de que muchos máximos que pueden ser candidatos a extremos no son considerados en este análisis, ya que sólo nos quedamos con el máximo de cada bloque lo que puede suponer una pérdida de información. Cuando se estudian los extremos por medio del método de bloques, es necesario tener un número de máximos total elevado, así como un número considerable de bins, y asegurarnos de que el número de cuentas por bin sea estadísticamente significativo, ya que el error de un bin en el que hay n cuentas es de $1/\sqrt{n}$. Estos aspectos son clave para lograr una buena estadística, pues la resolución del histograma dependerá de ellos.

2.1.1 Estimación de los parámetros GEV

Para estimar los parámetros de las distribuciones se hará uso de la función que dispone Matlab *Gevfit* que nos proporcionará el parámetro de forma ξ , el parámetro de escala σ y el parámetro de localización μ , con un intervalo de confianza del 95%. Eso significa que las estimaciones de los

parámetros que obtendremos resultarán ser el valor medio dentro de un intervalo lo suficientemente pequeño en donde la probabilidad de que dichos valores sean correctos es muy alta.

Existen varias técnicas para realizar este tipo de estimaciones, aunque la más común es el método de estimación por máxima verosimilitud y es la que precisamente utiliza Matlab. La función de verosimilitud para n variables aleatorias viene dada por la función de densidad de probabilidad conjunta de dichas variables dependiente de los tres parámetros:

$$L(x; \xi, \sigma, \mu) = \prod_{i=1}^n f(x_i) \quad (7)$$

Por sencillez, se utiliza el logaritmo de la función de verosimilitud:

$$\ln L [(x; \xi, \sigma, \mu)] = \sum_{i=1}^n \ln [f(x_i)] \quad (8)$$

Aplicado a la distribución de valores extremos generalizada, se tienen dos ecuaciones distintas: la primera para el caso en el que $\xi \neq 0$ (distribuciones Weibull y Fréchet) y la segunda para el caso en el que $\xi = 0$ (distribución de Gumbel).

$$\ln [L(x; \xi, \sigma, \mu)] = -n \ln \sigma - (1 + \frac{1}{\xi}) \sum_{i=1}^n \ln [1 + \xi (\frac{x_i - \mu}{\sigma})] - \sum_{i=1}^n [1 + \xi (\frac{x_i - \mu}{\sigma})]^{-\frac{1}{\xi}} \quad (9)$$

$$\ln [L(x; \sigma, \mu)] = -n \ln \sigma - \sum_{i=1}^n (\frac{x_i - \mu}{\sigma}) - \sum_{i=1}^n \exp [- (\frac{x_i - \mu}{\sigma})] \quad (10)$$

Donde para ambas se cumple que:

$$1 + \xi (\frac{x_i - \mu}{\sigma}) > 0 \\ i = 1, \dots, n.$$

Como estas ecuaciones no tienen solución analítica pero sin embargo disponemos de los datos que hemos generado, se pueden utilizar algoritmos de optimización como por ejemplo el test Kolmogórov-Smirnov, en donde en primer lugar computará μ y a continuación σ , para que finalmente, conociendo la distancia entre estos dos momentos pueda llevar a cabo una comparación con otra GEV teórica y así obtener ξ . Por tanto, los tres parámetros que maximicen las ecuaciones se corresponderán con los estimadores de máxima verosimilitud de la distribución de probabilidad.

2.2 Peak-over-threshold

Otra visión alternativa al tratamiento de los extremos sugiere fijar un umbral para así poder extraer de la serie original todos los máximos de interés definidos como los máximos de la variable en cada una de las excursiones sobre el umbral predefinido. La principal ventaja de este método es que, definiendo un umbral lo suficientemente alto, garantiza el estudio de los verdaderos extremos ya que puede haber máximos que no sean estrictamente extremos y por tanto no sean de interés. Además permite controlar el tamaño de la muestra aunque ésta ha de ser lo suficientemente larga para poder realizar el ajuste. Ello hace que sea el método predilecto para el análisis de extremos en series temporales, ya que hace un uso más eficiente de los datos contenidos en la serie inicial.

El teorema Pickands–Balkema–de Haan (1975) constituye el segundo teorema fundamental de la teoría de valores extremos clásica y viene a decir que las excedencias que superan un cierto umbral siguen la distribución generalizada de Pareto, dependiente de tres parámetros: forma, escala y localización.

Sin embargo, la dificultad que concierne a este método se centra en la elección de un umbral óptimo, que ha de ser escogido por el analista en función de la serie de datos de la que disponga. El único requisito que ha de cumplir es que sea lo suficientemente alto, ya que de este modo el comportamiento límite de las excedencias vendrá mejor determinado por el teorema. La función de distribución acumulada es de la forma siguiente:

$$GPD_{\xi}(x) = \begin{cases} 1 - (1 + \xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-\frac{x-\mu}{\sigma}) & \xi = 0. \end{cases} \quad (11)$$

y derivando, tenemos la densidad de probabilidad:

$$gpd_{\xi}(x) = \begin{cases} \frac{1}{\sigma} (1 + \xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}-1} & \xi \neq 0 \\ \frac{1}{\sigma} \exp(-\frac{x-\mu}{\sigma}) & \xi = 0. \end{cases} \quad (12)$$

Donde en los casos en que $\xi \geq 0$ se ha de cumplir que $x \geq \mu$, mientras que en el caso en el que $\xi < 0$ se necesitará que $\mu \geq x \geq \mu - \frac{\sigma}{\xi}$.

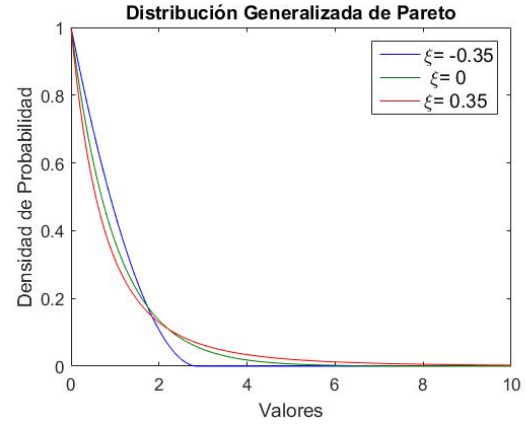


Figure 7: Familia de distribuciones de Pareto en función de ξ , para $\sigma = 1$ y $\mu = 0$. Notar que en este caso, cuando el parámetro de forma es nulo, la GPD_{ξ} se reduce a una exponencial.

Es importante mencionar que dada una serie temporal, se pueden obtener distribuciones de probabilidad diferentes utilizando ambos procedimientos. Esto se debe a la propia naturaleza de cada método ya que la selección de los eventos extremos no es la misma en ambos casos. Sin embargo si existen propiedades universales que son independientes de como se elijan los extremos. Concretamente, si tenemos las propiedades de los extremos para la metodología de bloques podemos obtener las propiedades de los extremos para el método de las excedencias y vice versa, ello puede hacerse por medio de la distribución de Pareto generalizada, cuyo parámetro de forma ξ coincide con el parámetro de forma de las distribuciones generalizadas de valores extremos.

$$1 + \ln(GEV_{\xi}(x)) = GPD_{\xi}(x) \quad (13)$$

Con lo cual, se puede decir, que gracias a la universalidad, la dificultad de predicción radicaría en obtener un buen ajuste que permita determinar la probabilidad de ocurrencia de eventos incluso más extremos que los registrados anteriormente ya que la precisión vendrá determinada por la calidad del ajuste.

Sin embargo, cabe destacar, que cuando los datos de la serie se encuentran fuertemente correlacionados las estimaciones a través del método de bloques y del método de las excedencias serán diferentes y la relación matemática que los hace equivalentes dejará de ser válida, lo que en la práctica conlleva a una mala estimación de los tiempos de retorno para los eventos extremos, es decir, el tiempo medio entre dos sucesos de baja probabilidad de ocurrencia.

En un principio, la teoría clásica de valores extremos fue desarrollada utilizando la metodología de los bloques, donde en el límite idealmente el número de bloques tiende a infinito, es decir, se considera que las series con las que se trabaja son muy largas y esa es la razón de que en este trabajo se lleve a cabo la aproximación por bloques, ya que en nuestro caso podremos generar series tan largas como queramos.

No obstante, en la realidad las series no tienen la longitud que idealmente gustaría, consecuentemente, el método de las excedencias y por consiguiente, la distribución generalizada de Pareto surge como un intento de adaptar la ideal teoría clásica de valores extremos a las series reales cotidianas, debido a esta limitación práctica y junto con el hecho de que la mayoría de procesos en la naturaleza presentan correlaciones de largo alcance, el método de preferencia generalmente es el método de las excedencias sobre un umbral.

3 Distribución tipo I: Gumbel

En primer lugar, vamos a ver cómo la distribución límite de los máximos para una distribución normal se ajusta a una distribución de Gumbel. Comenzaremos demostrando que la distribución

normal pertenece al dominio de atracción de la distribución de tipo I de los valores extremos. Para ello, utilizamos la condición necesaria y suficiente de Von Mises, que nos asegurará la distribución hacia la que convergerán los máximos:

$$\lim_{h \rightarrow x_f} \frac{1 - F(h + xg(h))}{1 - F(h)} = e^{-x} \quad \forall x$$

Donde $g(h)$ es una función real positiva y $x_f = \text{Sup}\{x : F(x) < 1\}$. Como x_f se corresponde con el punto derecho final de la distribución, tendremos que en este caso $x_f = \infty$. Aplicando el criterio tenemos:

$$\lim_{h \rightarrow \infty} \frac{1 - F(h + xg(h))}{1 - F(h)} \approx \lim_{h \rightarrow \infty} \frac{he^{-(h+xg)^2/2}}{e^{-h^2/2}(h+xg)} = \lim_{h \rightarrow \infty} \frac{e^{-\frac{(xg)^2}{2} - xgh}}{1 + \frac{xg}{h}} = e^{-x}$$

Si la función $g(h) = \frac{1}{h}$, entonces tendremos convergencia y el criterio se cumplirá.

El siguiente paso es comprobar numéricamente que esto se cumple, para ello, generaremos series a partir de números normales, los cuales serán generados a su vez, por la función que dispone Matlab *normrnd*, a pesar de que los números generados a través de esta rutina no son del todo aleatorios, si pueden ser considerados como tales, ya que solo se detecta que siguen un algoritmo de generación determinista tras un número de iteraciones que es del orden de 10^{23} .

A continuación, aplicaremos la metodología de los bloques, dividiendo la serie temporal de longitud s en K bloques de longitud N ($s = NK$). La serie con la que contamos en un principio tendrá una longitud de 10^9 datos, el procedimiento consistirá en generar 10^6 veces, series de longitud 10^3 , extrayendo de cada una de estas series su valor máximo. Posteriormente, representamos el vector que contiene todos los máximos y haremos un ajuste a la distribución de valores extremos generalizada. Para finalizar, utilizaré la función *gevfit* para estimar los parámetros de la distribución, corroborando que el parámetro de forma $\xi \simeq 0$, como corresponde a la distribución de extremos tipo I o Gumbel.

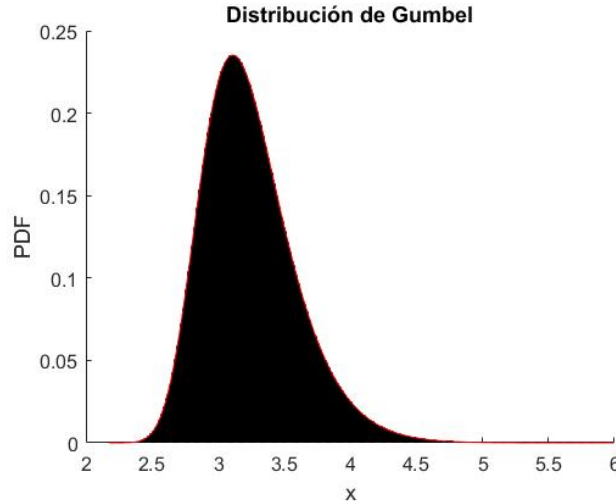


Figure 8: Distribución límite para los máximos de una distribución gaussiana. El método de bloques ha sido llevado a cabo considerando que el número de bloques es $K = 10^6$ con un tamaño de $N = 10^3$.

Como la distribución de Gumbel tiene un parámetro de forma nulo, podemos analizar la convergencia de dicho parámetro de ajuste en función del tamaño del bloque. Fijando el número de bloques, iremos variando la longitud de bloque observando, que a medida que el tamaño del bloque aumenta, el parámetro de forma toma valores cada vez más próximos a cero. Obsérvese que en el límite de tamaño de bloque $N \rightarrow \infty$ la distribución de extremos debe aproximarse a la distribución teórica de Gumbel.

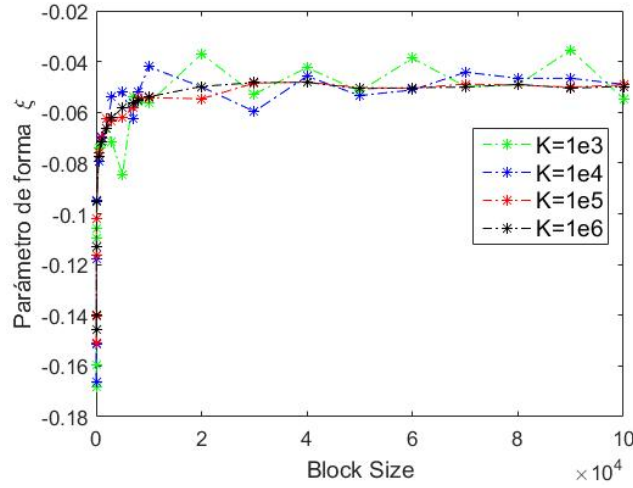


Figure 9: Convergencia del parámetro de forma para la distribución de Gumbel para una serie temporal gaussiana. Podemos observar como además del tamaño del bloque, también influye en la convergencia el número de bloques que consideremos, ya que para 10^6 bloques, se estabiliza mucho más rápido que para 10^4 bloques.

Además, no solo la convergencia del parámetro de forma se ve afectado por la variabilidad de la longitud del bloque, también los dos momentos restantes de la distribución, es decir, el parámetro de localización y el parámetro de escala, ambos pueden expresarse en función de sus valores medios para un número de máximos determinado y en función del tamaño del bloque N :

$$\mu = \overline{\mu_\infty} + \frac{1}{\sqrt{N}}$$

$$\sigma = \overline{\sigma_\infty} + \frac{1}{\sqrt{N}}$$

El error estadístico es gaussiano, ya que la diferencia entre cada valor obtenido para la escala y la localización respecto al valor medio de ambos, es decir, el error, es puramente aleatorio puesto que todas las series se han generado bajo las mismas condiciones. De las expresiones anteriores se desprende que conforme mayor sea el tamaño del bloque menor será el error estadístico y por tanto menor será la incertidumbre de los parámetros que caracterizan a la distribución.

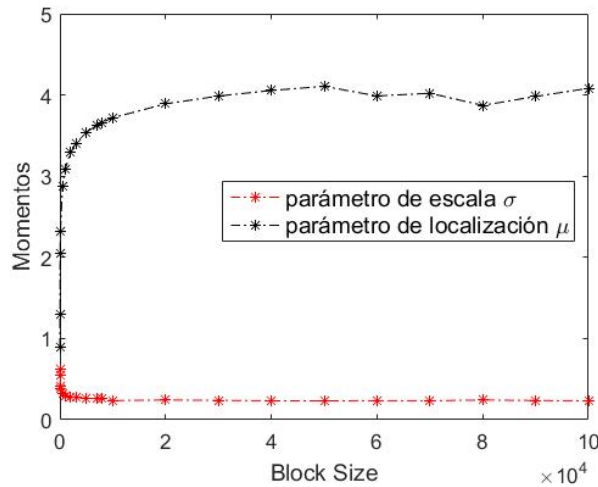


Figure 10: Convergencia del parámetro de escala y parámetro de localización en función de la longitud del bloque. La única restricción que ambos han de cumplir es que sean positivos.

4 Distribución tipo II: Fréchet

En esta sección estudiaremos la distribución de los eventos extremos en series temporales en las que la variable observada sigue una distribución con soporte semi-infinito. En particular generaremos series temporales a partir de la distribución $F(x) = 1 - x^{-1}$ si $x \geq 1$ y veremos como en este caso la PDF de los eventos extremos ya no se describe con una función Gumbel sino con una distribución de Fréchet. Para ello, comenzaremos demostrando que la distribución $F(x)$ pertenece al dominio de atracción de la distribución de tipo II de los valores extremos, utilizando la condición necesaria y suficiente de Von Mises:

$$\lim_{h \rightarrow \infty} \frac{1 - F(hx)}{1 - F(h)} = x^{-\alpha} \quad \forall x > 0, \alpha > 0.$$

Aplicando esta condición a nuestra distribución, tenemos que su complementaria es $\tilde{F}(x) = x^{-1}$ cuando $x \geq 1$ y por tanto vemos que se cumple:

$$\lim_{h \rightarrow \infty} \frac{\tilde{F}(hx)}{\tilde{F}(h)} = \lim_{h \rightarrow \infty} \frac{x^{-1}h^{-1}}{h^{-1}} = x^{-1} \quad x > 0, \alpha = 1 > 0.$$

En consecuencia, la distribución asintótica de los máximos se ajustará a una Fréchet. El siguiente paso es comprobar numericamente que este resultado se cumple, por lo que llevaremos a cabo el mismo procedimiento de antes, es decir, utilizaremos la metodología de los bloques particionando la serie temporal en K bloques de longitud N . Para ello, comenzaremos generando series de números aleatorios que sigan la distribución $F(x) = 1 - x^{-1}$ y tomaremos el máximo de cada una de las cajas de longitud N . Finalmente representaremos el vector de máximos obtenido y lo ajustaremos a la distribución generalizada de valores extremos.

Sin embargo, para poder generar las series que sigan dicha distribución $F(x)$ primero debemos de obtener el generador de números aleatorios, que será determinado aplicando el teorema de la función inversa a la distribución $F(x) = 1 - x^{-1}$:

$$F(x) = u$$

$$x = \frac{-1}{u - 1}$$

donde la variable transformada u sigue una distribución uniforme continua.

Una vez que ya se tiene el generador de números aleatorios para generar la series, aplicamos la metodología de los bloques y estimamos los parámetros de la distribución GEV utilizando la función *gevfit* de Matlab, observando que la distribución del máximo se ajusta a una distribución de Fréchet tal y como habíamos predicho.

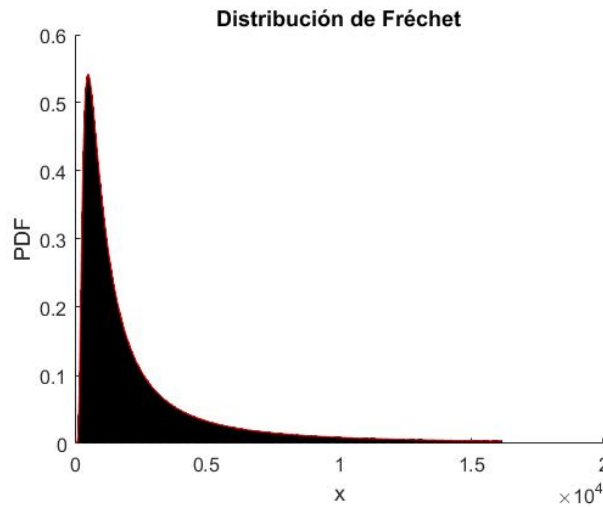


Figure 11: Cálculo de una PDF de los máximos de una serie temporal generada por la distribución $F(x) = 1 - x^{-1}$ si $x \geq 1$ y su ajuste a una distribución de Fréchet con parámetro de forma $\xi = 1.0061$ mediante el método de los bloques tomando un número de bloques $K = 10^6$ con un tamaño de bloque $N = 10^3$. Además vemos que en este caso, el máximo crece mucho más rápido que para la serie temporal gaussiana.

A diferencia de la distribución de Gumbel, el parámetro de forma para la distribución de Fréchet no tiene un valor definido, sino que puede tomar cualquier valor siempre y cuando éste sea positivo. Podemos observar cómo converge el parámetro de forma conforme variamos el tamaño del bloque. Como se puede apreciar en la figura 12 que se muestra a continuación, cuanto más grande sea el tamaño del bloque mejor será la convergencia, y en este caso, más se acercará a la unidad.

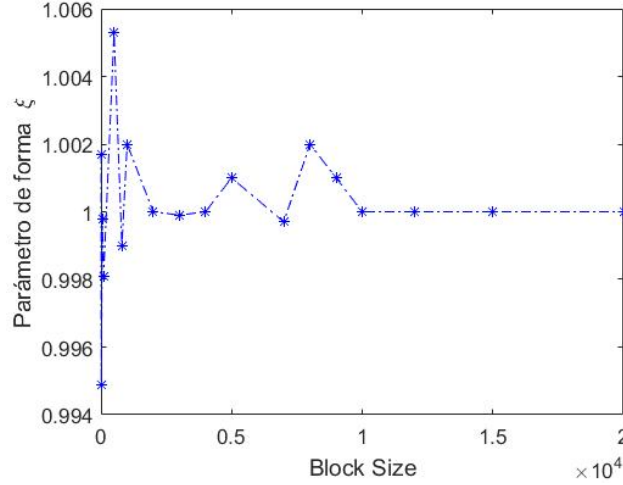


Figure 12: Convergencia del parámetro de forma para la serie temporal de la figura 11 hacia la distribución de Fréchet. Se observa que para un tamaño de bloque $N \sim 10^3$, ξ se estabiliza.

5 Distribución tipo III: Weibull

Finalmente, haciendo uso de la metodología de los bloques obtendremos numéricamente que la distribución de los valores extremos de una distribución uniforme se ajusta a una Weibull, como corresponde a series temporales aleatorias con una distribución con soporte acotado a un cierto intervalo. Para generar las series, utilizaremos números aleatorios acotados, comprendidos entre cero y uno, asociados a una distribución uniforme, para ello haremos uso de la rutina que ofrece Matlab *unifrnd*.

Desde un punto de vista teórico podemos esperar este resultado, ya que la distribución uniforme pertenece al dominio de atracción de la distribución de tipo III de los valores extremos. Esto puede comprobarse una vez más mediante la condición de Von Mises:

$$\lim_{h \rightarrow 0} \frac{1 - F(x_f - xh)}{1 - F(x_f - h)} = x^\alpha \quad \forall x > 0, \alpha > 0.$$

Dado que en este caso, la función de distribución acumulada complementaria viene dada por $\tilde{F}(x) = 1 - x$, tenemos que para el punto derecho final de la distribución $x_f = 1$, ya que el máximo convergerá a este valor cuando $n \rightarrow \infty$. Aplicando la condición, se tiene:

$$\lim_{h \rightarrow 0} \frac{\tilde{F}(1 - xh)}{\tilde{F}(1 - h)} = x \quad \forall x > 0, \alpha = 1 > 0. \quad (14)$$

Por tanto, los máximos seguirán una distribución de Weibull.

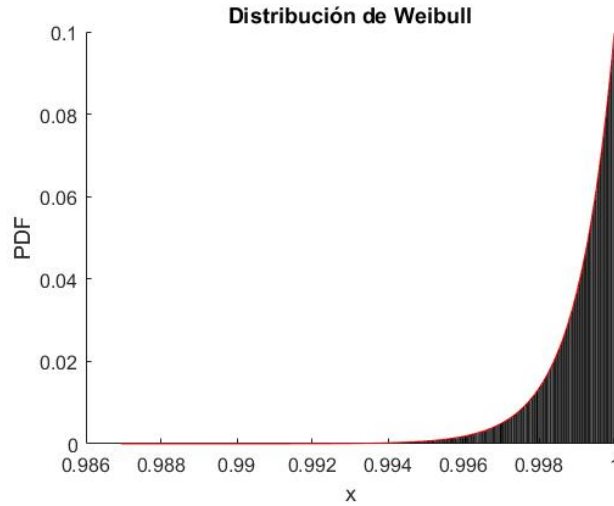


Figure 13: Distribución de los máximos de una serie temporal uniforme con un soporte acotado que ajusta a una distribución de valores extremos generalizada con un parámetro de forma $\xi = -0.9829$. El método de bloques ha sido llevado a cabo considerando el número de bloques $K = 10^6$ con una longitud de $N = 10^3$.

Al igual que la distribución de Fréchet, el parámetro de forma para la distribución de Weibull no toma un valor en particular, sino que en este caso, puede tomar cualquier valor siempre y cuando este sea negativo. Por tanto, si observamos qué valores para el parámetro de forma obtenemos al variar el tamaño del bloque, volvemos a ver que para tamaños de bloque grandes el parámetro de forma se estabiliza entorno a -1, mientras que para bloques pequeños toma valores más alejados y la incertidumbre de los parámetros es mayor.

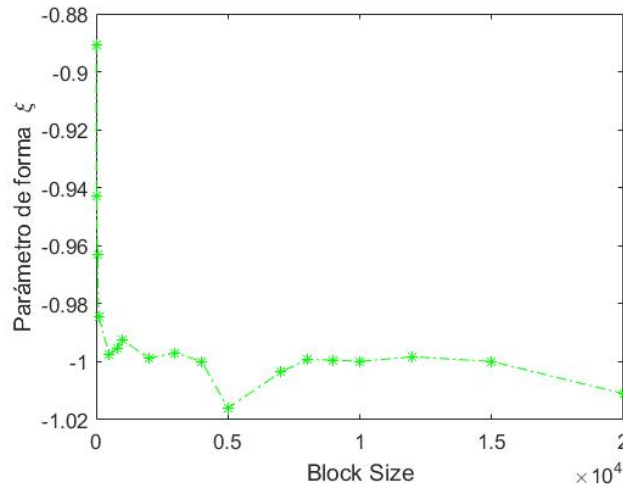


Figure 14: Convergencia del parámetro de forma para la distribución de Weibull. Observamos que para un tamaño de bloque $\sim 50^3$, ξ comienza a estabilizarse.

6 Efectos de la correlación temporal

La correlación es una manera de medir la dependencia entre las variables temporales de un proceso estocástico dependiente del tiempo. Se define la probabilidad a dos puntos como la probabilidad conjunta de tener un valor x_1 en t_1 y un valor x_2 en t_2 : Si dicha probabilidad conjunta puede factorizarse, es decir, $P(x_1, t_1, x_2, t_2) = P(x_1, t_1) \cdot P(x_2, t_2)$, tendremos que x_1 y x_2 son independientes y por tanto estarán descorrelacionados. Si por el contrario la probabilidad conjunta a dos puntos no puede ser factorizada, entonces ello implicaría que el proceso tiene memoria y existiría una dependencia estadística entre x_1 y x_2 .

La función de autocorrelación describe la correlación o la dependencia que existe entre dos valores en t y $t + \tau$, respectivamente para la misma variable aleatoria en una serie temporal, es decir, nos indica cómo se parecen dos observaciones de la misma variable en función del tiempo que transcurre entre ambas observaciones. Donde τ es el tiempo asociado a la distancia que hay entre los dos puntos de medida y se denomina *lag*.

$$C(\tau) = (x_t - \bar{x})(x_{t+\tau} - \bar{x})$$

Saber qué forma tiene esta función de correlación permite conocer si los datos de una serie siguen un determinado patrón, lo que supone una ventaja de cara a la predicción de eventos futuros, ya que de esta manera podremos saber si el sistema tiene poca ó mucha memoria, en función de su decaimiento y por ende, sabremos si será necesario estudiar series muy largas o con analizar los datos más recientes será suficiente para llevar a cabo predicciones. Además la función de correlación puede utilizarse para conocer la distancia temporal necesaria para que dos puntos de una serie dejen de estar correlacionados.

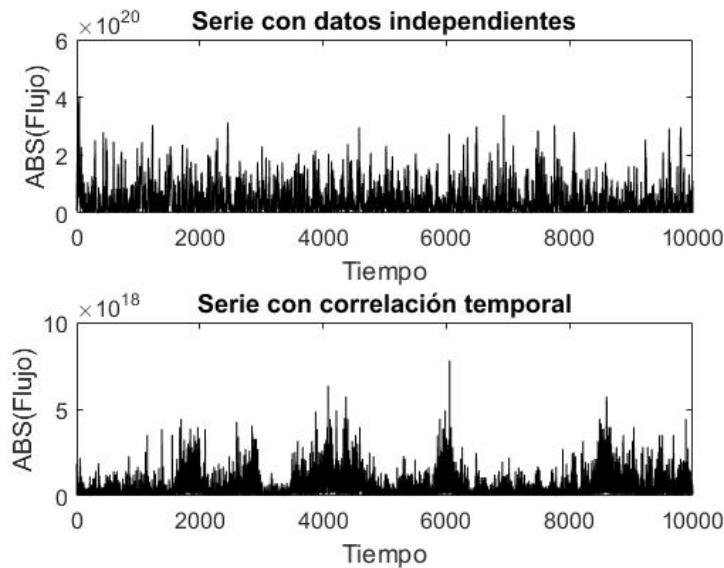


Figure 15: Ejemplo del efecto de la correlación temporal en una serie real procedente de un experimento llevado a cabo por la Universidad de Cantabria. En dicho experimento, consistente en el estudio de las correlaciones temporales en un plasma no confinado en una simetría cilíndrica, se muestra que para el centro del tubo el transporte de partículas es difusivo y por tanto los datos son independientes, mientras que para las regiones cercanas a los bordes del cilindro, existen correlaciones temporales de largo alcance y por tanto dependencia. Para más información, consultar referencia [16].

La teoría de valores extremos que hemos desarrollado en secciones precedentes nace de la suposición de que los datos de las series son independientes y por tanto estacionarios, una asunción que se ajusta a múltiples situaciones dado que todos los sistemas en el equilibrio son estacionarios, ahora bien, esto no siempre es así, un ejemplo son las glaciaciones, donde es necesario considerar grandes periodos de tiempo (escala de tiempo geológico) para poder hacer una descripción de este tipo de fenómenos. Las glaciaciones se dividen en periodos glaciares, para los cuales se estima un tiempo de retorno de entorno el orden de 10^5 años y es un claro ejemplo de en donde las correlaciones de largo alcance están presentes. Otros ejemplos abundan en la naturaleza e incluyen las series climáticas, las fluctuaciones de los valores de la bolsa o los fenómenos turbulentos entre otros muchos ejemplos.

En esta sección, veremos qué es lo que ocurre cuando introducimos correlación y hasta que punto esta es capaz de comprometer los resultados que nos proporciona la teoría de valores extremos clásica. Para ello, comenzaremos generando las series a partir de un sistema dinámico discreto que muestre un comportamiento caótico y por tanto memoria, finalmente, generaremos

series temporales a partir de un sistema dinámico continuo que vendrá definido por dos ecuaciones diferenciales estocásticas y que nos permitirán generar ruido correlacionado.

6.1 Series temporales con caos determinista

6.1.1 El mapa logístico

El sistema dinámico que utilizaremos será el mapa logístico que es un mapa unidimensional que viene descrito por la ecuación logística:

$$x_{n+1} = \mu x_n(1 - x_n) \quad x_n \in [0, 1], \quad n=0,1,2,\dots \quad (15)$$

La ecuación logística transforma un punto cualquiera en el intervalo unidad en otro punto dentro del mismo intervalo y su comportamiento vendrá descrito por el parámetro de control μ , definido para los valores $0 \leq \mu \leq 4$ y la condición inicial x_0 , que ha de ser tal que $0 \leq x_0 \leq 1$.

Para una condición inicial determinada x_0 , se procede a iterar la ecuación, generando un valor x_1 a partir de la relación $x_1 = \mu x_0(1 - x_0)$, nuevamente se genera x_2 , partiendo de x_1 , según $x_2 = \mu x_1(1 - x_1)$ y así sucesivamente. De esta manera obtendremos un conjunto de valores $\{x_0, x_1, x_2, \dots\}$ que es lo que se denomina la órbita del sistema dinámico.

Mediante este proceso iterativo se pueden encontrar los puntos fijos de la ecuación logística, que serán aquellos que cumplan que $x_n = x_{n+1}$ y que nos indican los valores hacia los que tiende la órbita del sistema, por lo que resolviendo la ecuación, los primeros puntos fijos que se encuentran son $x_1^* = 0$ y $x_2^* = \frac{\mu-1}{\mu}$, siendo el punto fijo $x_1^* = 0$ independiente de la condición inicial. Al aumentar el parámetro de control, aumentará el número de puntos fijos llegando a existir un número infinito de puntos fijos en la región caótica, este hecho puede apreciarse claramente en el diagrama de bifurcaciones.

El diagrama de bifurcaciones, también conocido como diagrama de Feigenbaum se obtiene cuando representamos todo el espectro del parámetro de control μ frente a la órbita logística para una condición inicial dada:

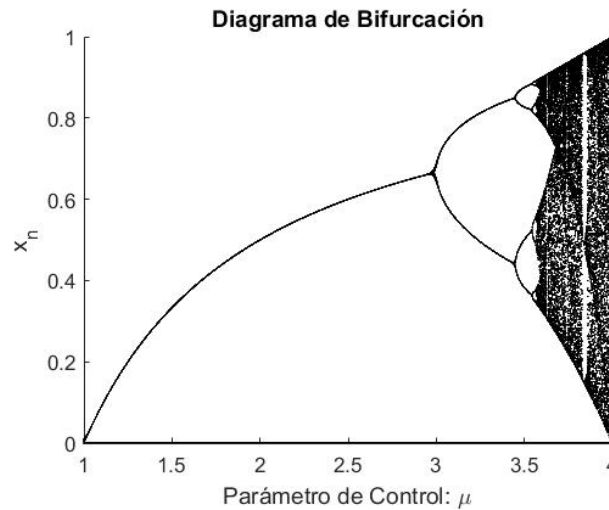


Figure 16: Diagrama de bifurcaciones de la ecuación logística para la condición inicial $x_0 = 0.1$. En general, se observa que cuando $0 < \mu < 3$ la órbita tiende a un cierto valor, para después producirse duplicaciones del periodo y finalmente, para $3.57 \leq \mu \leq 4.0$ presentar una dinámica caótica, donde de vez en cuando aparecerán ventanas de periodicidad.

Para $\mu < 1$ la órbita tenderá a cero independientemente de la condición inicial que se establezca y por eso la representación que hemos hecho comienza en $\mu = 1.0$. Si fijamos $x_0 = 0.1$ observamos que cuando $\mu = 2.0$ la órbita converge a 0.5. Sin embargo cuando el parámetro de control toma un valor de entorno a 3.0 se produce una bifurcación donde los valores que toman los x_n irán oscilando entre 0.4450 y 0.8419. Se observa además que conforme aumentamos el parámetro de control, para cada una de las ramas se vuelve a producir otra bifurcación, por ejemplo, podemos ver que cuando $\mu \approx 3.45$ se produce un ciclo de periodo cuatro y cuando $\mu \approx 3.55$ se produce un ciclo de periodo

ocho. Esta duplicación del periodo constituye una de las rutas hacia el caos y de hecho cuando $\mu \approx 3.57$ los valores x_n comenzarán a distribuirse de una manera aparentemente aleatoria dentro del intervalo unidad.

Conviene mencionar, que en la región caótica los valores de x_n en realidad no son aleatorios ya que la ecuación logística no contiene términos estocásticos. Las órbitas son deterministas, en el sentido de que conocemos la ecuación por la que se rigen, lo que quiere decir que si volviésemos a realizar el mismo número de iteraciones con la misma condición inicial y mismo parámetro de control, volveríamos a obtener los mismos valores de x_n , esto es lo que se conoce como caos determinista.

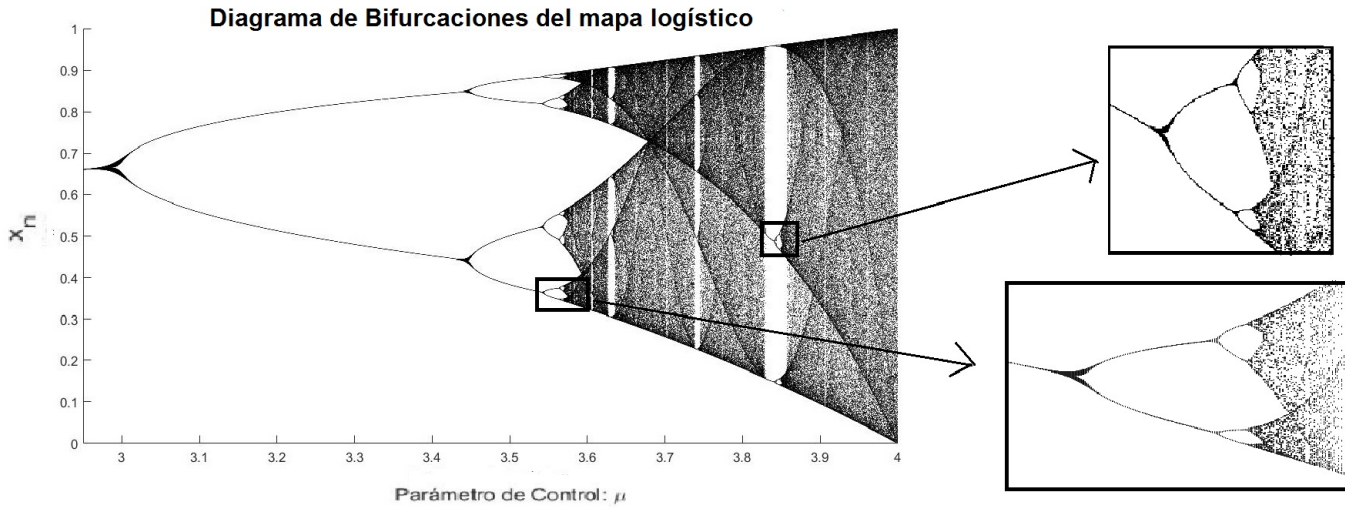


Figure 17: Ampliación de la figura 16 donde se muestra con mayor detalle cómo las duplicaciones del periodo conducen al caos, pudiéndose observar como las sucesivas bifurcaciones que aparecen son autosemejantes, por lo que la ecuación logística tiene un comportamiento fractal.

El mapa logístico presenta un comportamiento caótico, ya que para la región en la que $3.57 \leq \mu \leq 4$ se cumplen las condiciones mínimas que debe tener un sistema para que exista un comportamiento de este tipo: no tiene una conducta periódica, existe una sensibilidad a las condiciones iniciales y además posee una estructura fractal, es decir, bajo ampliación se verifica la característica de auto similitud.

Si aplicamos la metodología de los bloques al mapa logístico para estudiar sus extremos, esperamos obtener como distribución límite de los máximos, una distribución de Weibull, ya que como se ha visto, la ecuación logística está definida en un soporte acotado $[0,1]$ y por tanto, pertenecerá al dominio de atracción de la distribución de tipo III de los valores extremos. Puede comprobarse que se cumple la condición de Von Mises para la distribución de Weibull:

$$\lim_{h \rightarrow 0} \frac{1 - F(1 - xh)}{1 - F(1 - h)} = \lim_{h \rightarrow 0} \frac{\mu(1 - xh)xh}{\mu(1 - h)h} = \lim_{h \rightarrow 0} \frac{x(1 - xh)}{(1 - h)} = x \quad x > 0, \alpha = 1 > 0. \quad (16)$$

Sin embargo, como es de esperar, fuera de la región caótica no tiene sentido aplicar el modelo de bloques, ya que no habrá extremos, puesto que es la región periódica y por tanto solo veremos un pico correspondiente al único máximo que se alcanza:

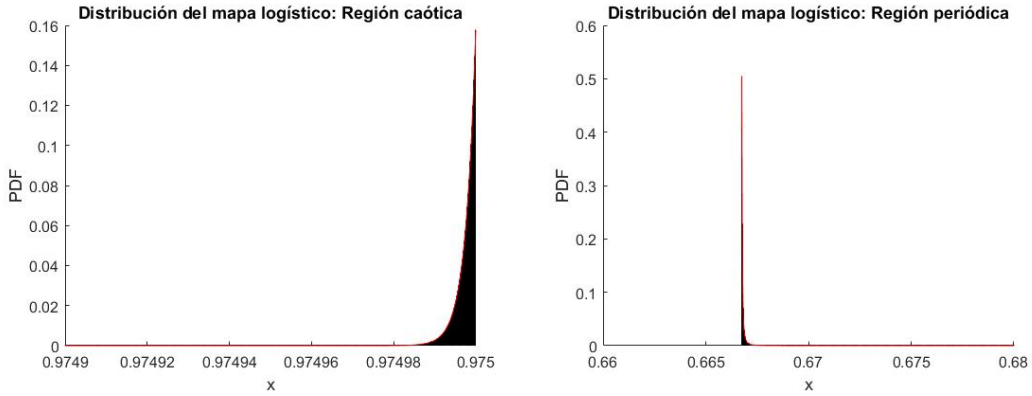


Figure 18: Aplicación del método de los bloques a las dos regiones del mapa logístico a partir de la condición inicial $x_0 = 0.1$. Restringiremos el estudio de los extremos a la región caótica. Las figuras han sido obtenidas considerando $\mu = 3.9$ y $\mu = 3.0$ respectivamente.

6.1.2 Exponentes de Lyapunov

Una cantidad que está íntimamente relacionada con la predictibilidad de los sistemas dinámicos son los exponentes de Lyapunov. Los exponentes de Lyapunov permiten cuantificar la velocidad con la que dos trayectorias que parten de una condición inicial muy próxima se acercan o se alejan conforme pasa el tiempo, lo que quiere decir que dan una medida del grado de sensibilidad del sistema a las condiciones iniciales.

Si partimos de dos condiciones iniciales para el instante inicial x_0 y $x_0 + \epsilon$, donde ϵ puede ser tan pequeño como uno quiera y se deja que el sistema evolucione con el transcurso del tiempo, se obtienen dos trayectorias. La diferencia entre ambas, define una evolución exponencial:

$$\delta x_t = \epsilon e^{\lambda t}$$

Si el exponente λ es negativo, las trayectorias confluirán en una sola cuando $t \rightarrow \infty$. Sin embargo, cuando λ sea positivo, las trayectorias se alejarán exponencialmente la una de la otra. Los sistemas con dinámica caótica han de tener al menos un exponente de Lyapunov positivo.

El número de exponentes de Lyapunov que tendrá un sistema dinámico dependerá del número de grados de libertad que lo caracterice, aunque su comportamiento vendrá dado por el mayor de todos sus exponentes. En nuestro caso, para el sistema logístico tendremos un solo exponente de Lyapunov ya que estamos trabajando con un mapa unidimensional. Para calcular el exponente de Lyapunov recurrimos al método derivativo, utilizando las dos condiciones iniciales mencionadas con anterioridad y donde f^n se corresponde con el mapa iterado n veces:

$$f^n(x_0 + \epsilon) - f^n(x_0) = (x_0 + \epsilon - x_0)e^{n\lambda(x_0)}$$

$$\lim_{\epsilon \rightarrow 0} \frac{f^n(x_0 + \epsilon) - f^n(x_0)}{\epsilon} = e^{n\lambda(x_0)}$$

Desde una perspectiva matemática el exponente de Lyapunov se puede relacionar con la matriz jacobiana, dado que el exponente describe cómo se comportan los vectores en el espacio tangente al espacio de fases: $\frac{df^n(x_0)}{dx}|_{x_0} = e^{n\lambda(x_0)}$. Desde un punto de vista más físico, la matriz jacobiana nos describiría como un pequeño cambio en x_0 se propagaría hasta el punto final $f^n(x_0)$. Por tanto, el exponente de Lyapunov vendrá definido según:

$$\lambda(x_0) = \frac{1}{n} \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \ln(|f'(x_i)|)$$

Podemos representar los valores que toma el exponente de Lyapunov en el mapa logístico en función del parámetro de control cuando $3 \leq \mu \leq 4$ (figura 19).

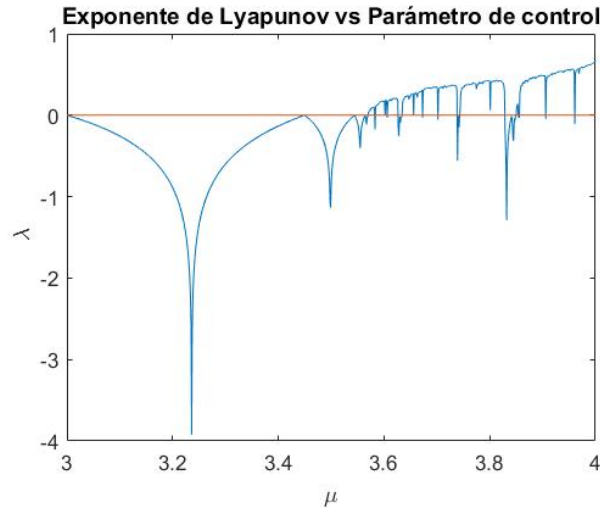


Figure 19: Relación entre el exponente de Lyapunov y el parámetro de control en el mapa logístico para $x_0 = 0.1$ y 1000 iteraciones. Para las regiones periódicas el exponente de Lyapunov es negativo, mientras que para las regiones caóticas el exponente de Lyapunov es positivo.

Podemos relacionar los resultados de esta figura con el diagrama de bifurcación mostrado anteriormente. Antes de llegar a la región caótica $3 \leq \mu < 3.57$, los exponentes de Lyapunov son negativos, los picos se corresponden con los ciclos de periodo dos, cuatro, ocho... respectivamente y observamos que cuando se va a producir una bifurcación, el exponente de Lyapunov tiende a cero. Para $3.57 \leq \mu \leq 4$, el exponente de Lyapunov cambia de signo y generalmente es positivo, lo que nos indica nuevamente la llegada del caos, vemos que progresivamente irá tomando valores crecientes, salvo en ciertas ocasiones donde decrece abruptamente, esto se debe a las ventanas periódicas que aparecen eventualmente en la región caótica.

Otra magnitud que se encuentra vinculada con los exponentes de Lyapunov es el llamado tiempo ó momento de Lyapunov cuya relación viene dada por:

$$\tau_c = \frac{1}{|\lambda|}$$

Este tiempo de Lyapunov es un tiempo característico de los sistemas dinámicos que permite estimar para qué período de tiempo se puede esperar que un sistema sea predecible. Si $\lambda < 0$ este tiempo característico funcionaría como un tiempo de relajación del atractor, ya que ambas órbitas se encontrarían muy próximas entre sí y tenderían a los mismos puntos fijos. Si por el contrario, $\lambda > 0$, sabemos que el sistema dinámico es caótico y este tiempo característico vendría a darnos el tiempo necesario para que ambas órbitas estén descorrelacionadas. En lo sucesivo, haremos referencia a este tiempo como tiempo de correlación.

6.1.3 Correlación de corto alcance

En los procesos donde hay correlaciones de corto alcance, la dependencia entre los valores de la variable aleatoria para diferentes tiempos disminuye rápidamente conforme aumenta el desfase temporal, es decir, dada una serie temporal, a sabiendas de que la última observación depende de todos los valores anteriores, observamos que esta dependencia será más fuerte con los datos más recientes y sin embargo, será más débil con los datos más alejados. En este caso, la función de correlación temporal sigue un decaimiento exponencial, donde τ_c es el tiempo de correlación:

$$C(\tau) \sim e^{-\frac{\tau}{\tau_c}}$$

Si el sistema tiene memoria a corto plazo, se tendrá que los tiempos de correlación son pequeños. Podemos estimar el tiempo de correlación sabiendo que $\tau_c \sim \frac{1}{\lambda}$, por lo que si por ejemplo en el caso en el que $\mu = 4$, el exponente de Lyapunov es $\lambda = \ln 2$, sea cual sea la condición inicial, por lo que el tiempo de correlación que se tiene es $\tau_c = 1.44$.

Fijando la condición inicial $x_0 = 0.1$ de cara a la aplicación del modelo de bloques, esperamos que la convergencia de la distribución no se vea afectada sustancialmente, al tratarse de un tiempo de correlación tan pequeño:

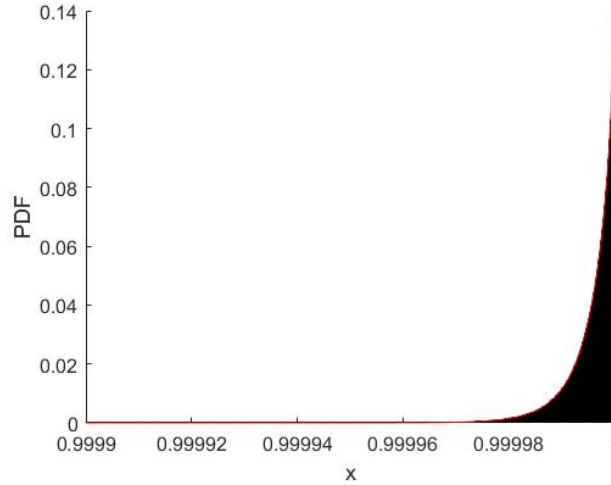


Figure 20: Distribución de valores extremos generalizada obtenida tras aplicar el método de los bloques a la ecuación logística considerando $\mu = 4$ ($\tau_c = 1.44$). El parámetro de forma de la distribución GEV es $\xi = -1.08$ correspondiente a una distribución de Weibull.

La teoría clásica de valores extremos apenas se altera cuando las variables aleatorias se encuentran débilmente correlacionadas y por tanto, puede aplicarse obteniendo unos resultados muy próximos a los que se obtendrían si las variables aleatorias fueran independientes, es decir, estuvieran descorrelacionadas.

6.1.4 Correlación de largo alcance

Existen otros procesos para los cuales la evolución del sistema está influenciada por los estados previos del sistema en tiempos mucho más grandes que el tiempo de correlación. En este caso, diremos que el sistema presenta memoria a largo plazo o lo que es lo mismo, correlaciones de largo alcance. En este tipo de procesos, la función de correlación temporal muestra un decaimiento de tipo algebraico, que resulta ser mucho más lento para tiempos de correlación largos, que en el caso exponencial anterior:

$$C(\tau) \sim \tau^{-\gamma} \quad \gamma > 0. \quad (17)$$

La correlación hace que la convergencia de la distribución sea más lenta, por lo que para tiempos de correlación largos tardaremos mucho más en alcanzar dicha convergencia que para tiempos cortos donde obtendremos un buen ajuste rápidamente. Podemos comprobar los efectos de las correlaciones de largo alcance aplicando el modelo de bloques nuevamente a la ecuación logística. Para ello comenzaremos explorando una región del mapa logístico donde el exponente de Lyapunov aún positivo, esté muy cercano a cero, ya que cuanto más pequeño sea éste, mayor será el tiempo de correlación y por tanto mayor será la dependencia a largo plazo del proceso ($\tau_c = 1/\lambda$).

Los exponentes de Lyapunov que cumplen este requisito se obtienen para parámetros de control comprendidos $3.57 < \mu < 3.92$. En este caso, se ha seleccionado para $\mu = 3.597$ un exponente de Lyapunov $\lambda = 0.02636$, de esta manera se obtiene un tiempo de correlación $\tau_c = 37.94 [T]$.

Para observar mejor los efectos de la correlación, utilizaremos dos tamaños de bloque distintos y veremos cómo modificando la escala en ambos casos se consigue hacer una aproximación de la independencia de los datos, para que de esta manera la teoría de valores extremos sea capaz de proporcionarnos resultados válidos.

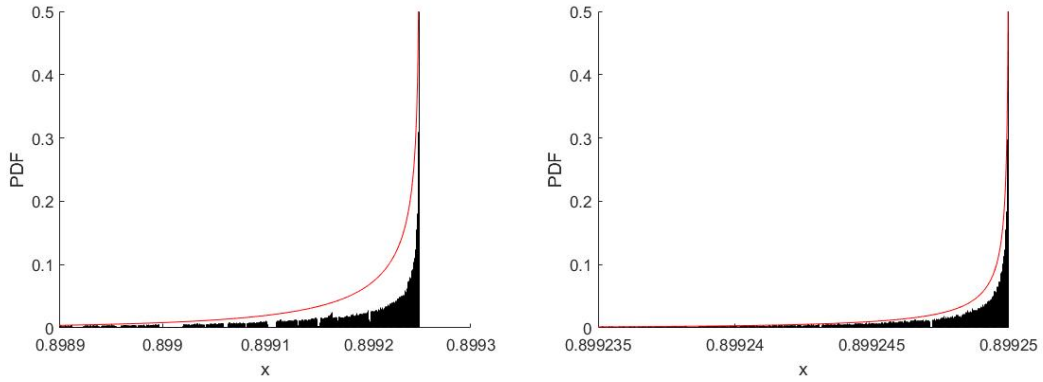


Figure 21: Distribuciones generalizadas de valores extremos generadas a partir de un mismo número de bloques $K = 10^6$ y con un tamaño de bloque de $N = 10^2$ y $N = 10^3$ respectivamente. A su vez, los parámetros de forma que se obtienen son $\xi = -1.65$ y $\xi = -1.43$, por lo que ambas distribuciones GEV se corresponden con una distribución de Weibull.

Bajo estas condiciones, para obtener un buen ajuste es preciso llevar a cabo una reescala del tamaño del bloque, para que de esta manera se puedan dejar de apreciar los efectos provocados por la correlación. La reescala se llevará a cabo de la siguiente manera:

$$N' = \frac{N}{\tau_c} \quad \text{Siendo } N \text{ el tamaño de bloque original y } N' \text{ el bloque final reescalado.} \quad (18)$$

En este caso, como $\tau_c = 37.94$, se tendrá que debemos de multiplicar al menos por un factor 38 el tamaño del bloque original, para que los datos se vayan ajustando cada vez más al límite teórico. Por tanto si comenzamos aumentando un factor 10 el tamaño de bloque, observamos que la convergencia es notablemente mejor:

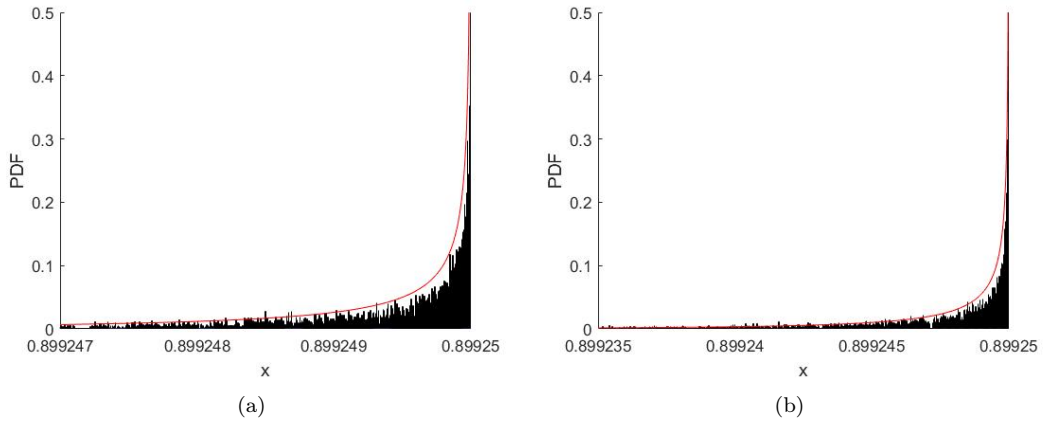


Figure 22: En la figura a) se muestra la distribución de valores extremos generalizada generada tras aumentar un factor 10 el tamaño de bloque inicial por lo que se ha utilizado un tamaño de $N = 10^3$ con $K = 10^6$ bloques, obteniendo un parámetro de forma $\xi = -1.86$ correspondiente a una distribución de Weibull. Idem para la figura b) pero utilizando un tamaño de bloque $N = 10^4$ obteniendo un parámetro de forma para la GEV $\xi = -1.74$ indicando que también se corresponde con una distribución de tipo III de los valores extremos.

Si finalmente, multiplicamos por un factor 40, observamos que la convergencia ya es del todo buena:

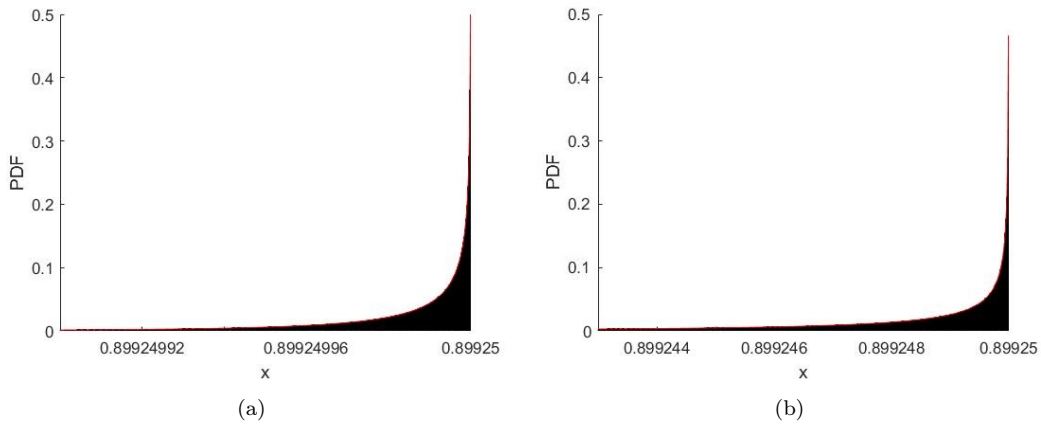


Figure 23: En la figura a) se muestra la distribución de valores extremos generalizada generada a partir de $K = 10^6$ bloques y tras aumentar un factor 40 el tamaño de bloque inicial por lo que se ha utilizado un tamaño de $N = 40^3$, obteniendo un parámetro de forma $\xi = -1.91$ asociado a una distribución de Weibull. Idem para la figura b) pero utilizando un tamaño de bloque de un orden de magnitud mayor a la anterior, $N = 40^4$ obteniendo un parámetro de forma para la GEV $\xi = -1.86$ indicando que también se corresponde con una distribución de tipo III de los valores extremos.

Por lo general, los tiempos de correlación que aparecen en las series reales son mucho mayores que el tiempo de correlación que hemos tratado, como por ejemplo, en el caso de los terremotos, que son fenómenos típicos que se caracterizan por tener dependencias a largo plazo, y en donde se observa que en tiempos previos al terremoto principal, aparecen otros terremotos de menor magnitud, cuya frecuencia se irá incrementando conforme aumente la tensión de la falla.

La predicción para procesos de larga memoria, esto es, cuando tengamos tiempos de correlación $\tau_c > 10$, podrá llevarse a cabo, realizando la reescala del bloque con el fin de que los datos aparezcan descorrelacionados, ya que la teoría de valores extremos exige que los eventos sean independientes, de esta manera, se podrá calcular su distribución y finalmente, obtener del histograma la probabilidad de ocurrencia de un extremo en cuestión. Una vez que se conoce la probabilidad del evento, podemos estimar el periodo de retorno, que es el tiempo medio entre dos sucesos raros, el cual no deja de ser un concepto estadístico pero que resulta de gran utilidad, ya que nos indica el intervalo de recurrencia promedio durante un periodo de tiempo prolongado. Pongamos un ejemplo:

Supongamos que estamos estudiando las inundaciones a partir de datos recopilados durante 1000 años. Obtenemos nuestra distribución, que sabremos que será una de las tres distribuciones asintóticas, a partir de ella, obtenemos que la probabilidad de que se vuelva a producir una inundación en un año es del 1%, lo que implicaría un tiempo de retorno de 100 años. Sin embargo, este tiempo de retorno no nos indica que el suceso se vaya a producir una vez cada 100 años, ó que se vaya a producir con una periodicidad de 100 años. Lo que realmente nos indica es que en un intervalo de tiempo de 100 años, el suceso podrá darse tantas veces como sea, o incluso puede que no se de nunca, teniendo cada resultado una probabilidad asociada, que podrá obtenerse a partir de la distribución de probabilidad de los tiempos de retorno.

Si los eventos son independientes o existe un tiempo de correlación pero éste es pequeño, los tiempos de retorno también estarán descorrelacionados y por tanto seguirán una distribución de Poisson, donde T se corresponde con el tiempo de retorno medio:

$$P(t) = \frac{1}{T} \exp\left(-\frac{t}{T}\right)$$

Si por el contrario, los eventos tienen memoria a largo plazo, los tiempos de retorno presentarán correlación de largo alcance y en este caso, seguirán una distribución del tipo:

$$P(t) \sim \exp\left[-cte \left(\frac{t}{T}\right)^\gamma\right] \quad 0 < \gamma < 1 \quad (19)$$

6.2 Series temporales estocásticas

En este apartado en lugar de considerar un sistema dinámico discreto, consideraremos un sistema dinámico continuo que vendrá dado por una ecuación diferencial con un término estocástico, de esta manera podremos trabajar con series temporales procedentes de un proceso físico cualquiera que varía aleatoriamente en el tiempo y cuyas variables estocásticas describen el estado del sistema en cada instante de tiempo.

Los procesos estocásticos tienen una gran relevancia en los sistemas complejos ya que suelen utilizarse para intentar modelar diversos fenómenos, ya sean físicos, económicos ó biológicos, puesto que en general suele existir una determinada variable o parámetro del sistema que presenta un comportamiento inestable y cuyas fluctuaciones impiden describir el comportamiento del sistema mediante una ecuación diferencial determinista. Bajo esta situación, resulta más conveniente definir sus variaciones en términos probabilísticos y estas probabilidades dependerán del estado actual y de los estados anteriores del sistema.

El proceso estocástico con el que trabajaremos será un proceso de difusión, conocido como proceso de Ornstein-Uhlenbeck que de forma genérica está compuesto por un sistema de dos ecuaciones:

$$\begin{aligned}\dot{x} &= f(x) + g(x)\Gamma(t) \\ \dot{\Gamma} &= -\frac{\Gamma}{\tau_c} + \frac{\zeta(t)}{\tau_c}\end{aligned}$$

siendo $\zeta(t)$ el ruido blanco gaussiano sin correlación.

Este proceso, fue propuesto por Ornstein y Uhlenbeck para tratar de modelar la variación de la velocidad de una partícula en un medio viscoso bajo una fuerza de fricción $f(x)$. Por simplicidad supondremos que la partícula se mueve con velocidad constante y sin fuerza de rozamiento por tanto, nuestra serie estará formada por una sucesión de variables aleatorias que se corresponderán con las posiciones de la partícula en el tiempo t . En lo sucesivo, trabajaremos considerando que $f(x) = 0$ y $g(x) = 1$, ya que el ruido de las series temporales que queremos generar queremos que sea aditivo.

Para obtener las series temporales a partir de dicho proceso debemos de tener en cuenta que contamos con una ecuación diferencial estocástica por lo que para generar las series primero debemos integrar ambas ecuaciones simultáneamente. Sin embargo en este caso no podemos recurrir a los métodos de integración clásicos para ecuaciones diferenciales ordinarias, ya que no podemos integrar el término aleatorio debido a que sus variaciones no están acotadas y a que no es diferenciable en ningún punto.

De la integración de estas ecuaciones utilizando el método de Euler-Maruyama se obtiene:

$$x(t + \Delta) = x(t) + f(x)\Delta + \Gamma(t)\Delta + O(\Delta^{\frac{3}{2}}) \quad (20)$$

$$\Gamma(t + \Delta) = \Gamma(t)(1 - \frac{\Delta}{\tau_c}) + \frac{X_1(t)}{\tau_c} + O(\Delta^{\frac{3}{2}}) \quad (21)$$

En primer lugar, obtendremos los valores extremos de la serie aplicando la metodología de los bloques y utilizando como algoritmo generador de las variables aleatorias, las dos ecuaciones que constituyen la solución de la ecuación diferencial estocástica. Para ello, comenzaremos definiendo un vector de tiempos cuya longitud total del vector se corresponderá con el tamaño de bloque N de la serie y que dividiremos en diferenciales de tiempo muy pequeños Δ , para que de esta manera el tiempo sea aproximadamente continuo. Además, necesitamos definir una condición inicial para el ruido que será tal que $\Gamma(0) = 1$, para que de esta manera podamos comenzar a calcular para cada valor de t contenido en el vector de tiempos, la contribución del ruido $\Gamma(t + \Delta)$, que introduciremos a su vez en la expresión $x(t + \Delta)$ y que nos proporcionará la sucesión de variables aleatorias correspondientes a la posición de la partícula en cada momento. El número de veces que se repite este proceso para todos los valores de t que componen el vector de tiempos, se corresponderá con el número de bloques K , y tomaremos de cada uno de ellos el valor máximo de $x(t + \Delta)$, obteniendo finalmente un vector de máximos que ajustaremos mediante el comando de matlab *gevfit* a una distribución de valores extremos, obteniendo así su parámetro de forma.

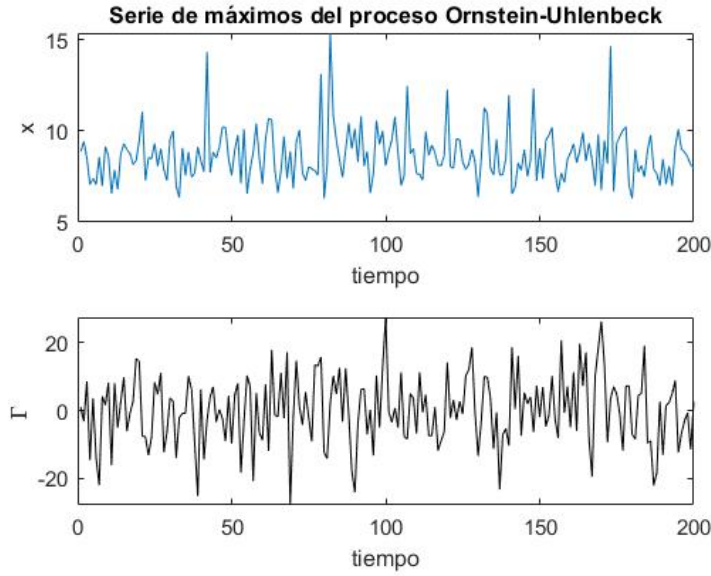


Figure 24: En la parte superior se muestra la trayectoria que siguen los máximos tras aplicar la metodología de los bloques al proceso estocástico difusivo de Ornstein-Uhlenbeck, que representa el movimiento de una partícula browniana en un fluido sin fuerza de fricción. En la parte inferior se muestra la contribución ruidosa del proceso que es de tipo gaussiano. Ambas representaciones se han realizado considerando un número de bloques $K = 10^4$ con una longitud de $N = 10^4$, un diferencial de tiempo $\Delta = 0.01$ y un tiempo de correlación $\tau_c = 0.1$

La componente aleatoria Γ del proceso de Ornstein-Uhlenbeck depende del tiempo de correlación τ_c y del ruido blanco ξ , es decir, un proceso estocástico que genera datos independientes donde no hay correlación estadística entre los distintos valores y su promedio es cero $\langle \xi \rangle = 0$, mientras que la covarianza estará siempre bien definida salvo cuando $t = t + \tau$ que será infinita, es por ello que la función de correlación tendrá la siguiente forma:

$$\langle \xi(t) \cdot \xi(t + \tau) \rangle = 2\sigma^2 \delta(\tau)$$

La razón por la que se ha introducido la ecuación diferencial para Γ es porque queremos generar ruido con correlaciones largas y es por ello que la función de correlación en este caso tendrá la siguiente forma:

$$\langle \Gamma(t) \cdot \Gamma(t + \tau) \rangle \sim e^{-\frac{\tau}{\tau_c}}$$

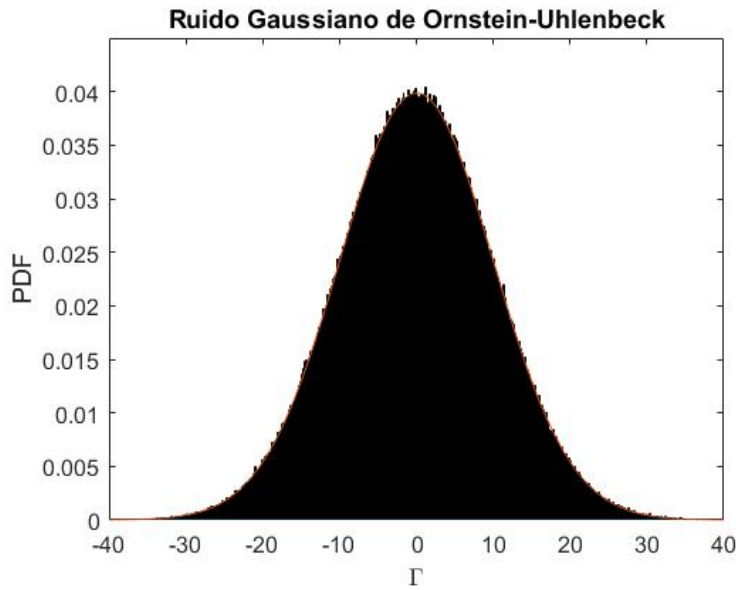


Figure 25: Los valores de la componente aleatoria del proceso de Ornstein-Uhlenbeck siguen una distribución normal. La figura se ha obtenido considerando una serie de tamaño $s = 10^5$, un $\tau_c = 0.1$ y una condición inicial $\Gamma(0) = 1$.

A continuación veremos cómo afecta la correlación temporal a este tipo de series, la principal ventaja de este proceso frente al sistema discreto logístico que hemos utilizado anteriormente, es que podremos generar series con correlación temporal a nuestra voluntad, es decir, podremos establecer el tiempo de correlación τ_c que queramos con tan solo modificar dicho valor en la ecuación (21), que describe el término estocástico del proceso de Ornstein-Uhlenbeck. En este caso, el tiempo de correlación τ_c nos da una medida de la rapidez con la que cambia la velocidad de la partícula.

En primer lugar, debemos de conocer a que dominio de atracción pertenece nuestro proceso estocástico para así poder saber a cual de las tres distribuciones asintóticas tenderán los máximos de la solución de la ecuación diferencial. Sin embargo, si tenemos en cuenta que el proceso de Ornstein-Uhlenbeck está constituido por una parte determinista y una parte aleatoria que sigue una distribución normal, hemos visto que aplicando la correspondiente condición de Von Mises a la distribución gaussiana, obtenemos que dicha distribución pertenece a la cuenca de atracción de la distribución de tipo I de los valores extremos o Gumbel (*ref: Sección 3*), por lo que podemos afirmar que la distribución límite de los máximos del proceso de Ornstein-Uhlenbeck será también una distribución de Gumbel ($\xi = 0$).

En segundo lugar, aplicaremos el modelo de bloques a la solución de nuestro proceso y una vez fijado un número de bloques $K = 10^4$, iremos variando el tamaño de bloque N para analizar la convergencia del parámetro de forma ξ , que deberá aproximarse a cero conforme el tamaño de bloque aumenta. No obstante, realizaremos este mismo procedimiento para distintos tiempos de correlación, observando cómo influye la magnitud del tiempo de correlación en la rapidez con la que converge el parámetro de forma. Para tiempos de correlación muy pequeños, la convergencia hacia el valor teórico del parámetro de forma ξ es bastante rápida tal y como es de esperar, dado que prácticamente hay independencia entre los datos y bajo esta premisa fundamental, la teoría de valores extremos se cumple sin problema alguno.

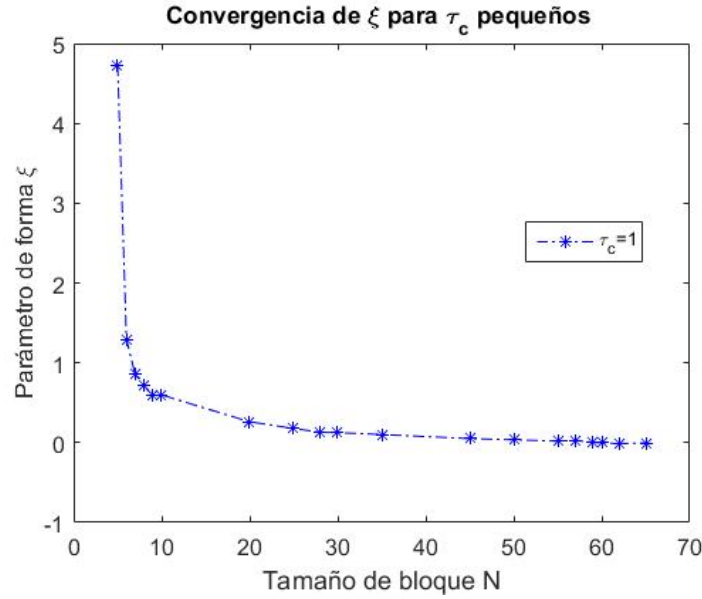


Figure 26: Ejemplo de cómo el parámetro de forma converge rápidamente a cero cuando introducimos a la serie temporal un tiempo de correlación muy pequeño ($\tau_c = 1$) de forma que $N \gg \tau_c$ y por tanto los datos de la serie temporal son aproximadamente independientes, teniendo un proceso de difusión con poca memoria.

Sin embargo cuando los tiempos de correlación son largos, surge la necesidad de aumentar notablemente el número de iteraciones tomando tamaños de bloque cada vez más grandes para empezar a observar la convergencia del parámetro de forma, ya que ésta se alcanza de una manera muchísimo más lenta que en el caso anterior.

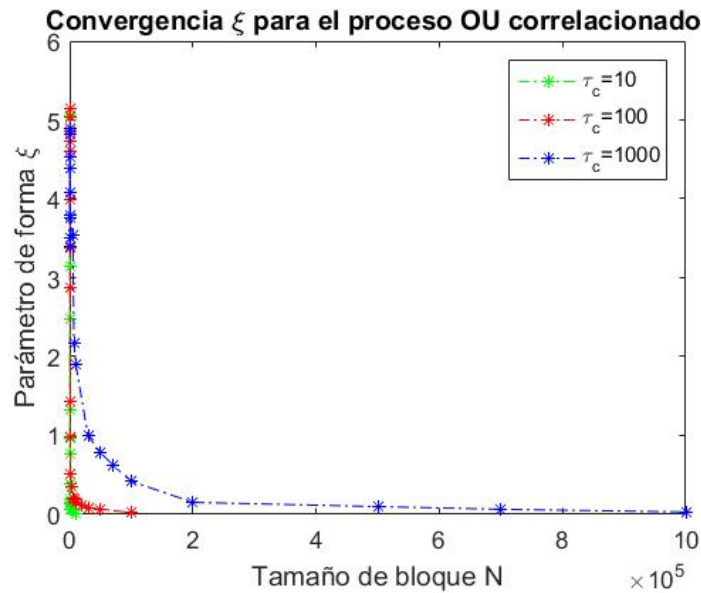


Figure 27: Las correlaciones largas implican $\tau_c \gg N$, por lo que para un tiempo de correlación medio como es el caso de $\tau_c = 10$, podemos ver que comenzaríamos a obtener convergencia de ξ a partir de $N \sim 2 \cdot 10^3$, por lo que necesitaríamos utilizar un tamaño de bloque de dicho orden de magnitud como mínimo para obtener parámetros de forma válidos. Para los siguientes tiempos de correlación que son más largos, el proceso está mucho más correlacionado y por tanto el orden de magnitud del tamaño de bloque que debemos de utilizar para el cual podemos comenzar a notar la convergencia de ξ sería de $N \sim 3 \cdot 10^4$ para el caso en el que $\tau_c = 100$ y de $N \sim 5 \cdot 10^5$ para el caso de $\tau_c = 1000$.

En la figura podemos apreciar para los tres casos que cuando el tiempo de correlación es mucho mayor que el tamaño de bloque N , la serie está mas correlacionada y por tanto, los valores que se obtienen para el parámetro de forma ξ difieren considerablemente del parámetro de forma teórico ($\xi = 0$). Por el contrario, observamos que cuando el tiempo de correlación comienza a ser menor que el tamaño de bloque, las correlaciones son pequeñas y de esta manera obtenemos ya valores para el parámetro de forma más adecuados.

Cabe destacar que llegar a identificar las clases de universalidad de la estadística de extremos de una serie de variables aleatorias fuertemente correlacionadas suele ser un problema, debido a que en ocasiones existe una limitación práctica como consecuencia de la necesidad de efectuar numerosas iteraciones utilizando un tamaño de bloque cada vez más grande ($N \rightarrow \infty$) por lo que hay veces que el tamaño de bloque que hay que utilizar es tan grande que no se puede llevar a cabo este procedimiento de renormalización ($N' = \frac{N}{\tau_c}$), y por tanto no se puede conocer el valor hacia el que tiende el parámetro de forma debido a que aún no se pudo alcanzar la convergencia de dicho parámetro, lo que imposibilita saber a que distribución límite se ajustan los extremos del proceso.

7 Conclusiones

Un evento extremo es un suceso de muy baja probabilidad pero que sin embargo, trae consigo consecuencias devastadoras. Dado que su probabilidad de ocurrencia es muy pequeña, este tipo de eventos estarán localizados en las colas de las distribuciones de probabilidad, por ello surge la necesidad de establecer un modelo estadístico construido en el marco de una teoría matemática relativa a los extremos, que centre toda su atención en las colas de las distribuciones. Este modelo, es la distribución de valores extremos generalizada, formada a partir de tres distribuciones que difieren en su parámetro de forma: Gumbel ($\xi = 0$), Fréchet ($\xi > 0$) y Weibull ($\xi < 0$), y que pone de manifiesto la universalidad de los extremos, estableciendo que los máximos de una serie tenderán a una de estas tres distribuciones asintóticas dependiendo del soporte de la distribución que siga la serie temporal original.

En primer lugar, hemos comprobado esta universalidad a partir de tres distribuciones propuestas y para ello hemos utilizado una de las técnicas que se emplean actualmente para estudiar los extremos, que es el método de bloques. Este método, consiste en dividir la serie temporal en bloques y extraer de cada uno de ellos el máximo, para finalmente hacer un ajuste a la distribución de valores extremos generalizada. A su vez, hemos comprobado cómo influye el tamaño de bloque en la convergencia del parámetro de forma, observando cómo cuando el tamaño de bloque es grande, la convergencia es mucho más rápida.

En segundo lugar, hemos introducido el mapa logístico, que es un sistema dinámico discreto, que nos permite generar series con correlación temporal. Para estudiar hasta qué punto afecta la correlación a los resultados que nos proporciona la teoría de valores extremos, consideramos una magnitud típica de los sistemas dinámicos que es el exponente de Lyapunov y que nos permite variar el tiempo de correlación, que nos da una idea de la distancia que debe de haber entre dos puntos para que se encuentren descorrelacionados. De esta manera, aplicando nuevamente la metodología de los bloques, hemos podido verificar cómo los resultados de la teoría de valores extremos apenas se ve afectada cuando los tiempos de correlación son pequeños, mientras que cuando tenemos tiempos de correlación largos, lo que implica que el sistema presenta una dependencia a largo plazo, los resultados no son tan buenos y entonces es preciso llevar a cabo una reescala del tamaño del bloque en función del tiempo de correlación, con la finalidad de que los datos se encuentren lo más alejados posible para aproximar independencia entre ellos.

En tercer lugar, hemos recurrido a un sistema de ecuaciones diferenciales estocásticas conocido como proceso de difusión de Ornstein-Uhlenbeck y que tras su integración, nos ha permitido generar series temporales estocásticas con ruido correlacionado. Dado que la componente aleatoria del proceso se ajusta a una distribución normal, independientemente del tiempo de correlación considerado, los máximos obtenidos de aplicar la metodología de los bloques al proceso de difusión tendrán una distribución límite de Gumbel, lo que nos ha permitido ver numéricamente cómo la rapidez de la convergencia se ve fuertemente afectada en función del tiempo de correlación que se considere.

Por tanto, bajo correlaciones de largo alcance, la convergencia de la distribución generalizada de valores extremos se alcanza más lentamente que si no hubiera correlación o si el sistema presentara memoria a corto plazo, ello es de esperar, ya que toda la teoría de valores extremos está construida a partir del supuesto de independencia entre las variables aleatorias que componen la serie.

8 Bibliografía

1. Lucarini, V., Faranda, D., Freitas, A.C.M., Milhazes, J., Holland, M., Kuna, T., Nicol, M., Todd, M., Vaienti, S. (2015). *Extremes and Recurrence in Dynamical Systems*. New York: Wiley.
2. Schweigler, T., Davidsen, J. (2011). Clustering of extreme and recurrent events in deterministic chaotic systems. Complexity Science Group, Department of Physics & Astronomy, University of Calgary. *Physical Review*.
3. Bovier, A. Extreme values of random processes. Institut für Angewandte Mathematik Wegelerstrasse 6, 53115 Bonn, Germany.
4. Bunde, A., Eichner, J. F., Kantelhardt, J. W., & Havlin, S. (2003) The effect of long-term correlations on the return periods of rare events. *Physica A* 330, 1-7.
5. Bunde, A., Eichner, J. F., Kantelhardt, J. W., & Havlin, S. (2004) Return intervals of rare events in records with long-term persistence. *Physica A* 342, 308-314.
6. Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions*. London: Imperial College Press.
7. Gumbel, E.J. (1954). *Statistical theory of extreme values and some practical applications*. Applied Mathematics Series 33. 1st edition. U.S. Department of Commerce, National Bureau of Standards.
8. Galambos, J. (1987). *The asymptotic theory of extreme order statistics*. 2nd edition, Malabar, FL: Robert E. Krieger.
9. Mcneil A, J. (1997). Estimating the Tails of loss severity distributions using extreme value theory. Departement Mathematik ETH Zentrum CH-8092 Zurich.
10. Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, Berlin.
11. Calculate lyapunov of the logistic map. (s.f.). Obtenido de:
<http://lab.fs.uni-lj.si/lasin/www/teaching/chaotic/Lyaplogistic/lylogistic.html>
12. IBM Knowledge Center. Transformaciones de series. (s.f.). Obtenido de:
https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_maihhelp_client_adita/components/dt/timeseries_transformations.html
13. Barceló, T. Departamento de Matemáticas de la Universidad Autónoma de Madrid. La ecuación logística (s.f.). Obtenido de:
http://www.uam.es/personal_pdi/ciencias/barcelo/cnumerico/recursos/logistica.htm
14. Peters, O Hertlein, C & Christensen, K. (2002). A Complexity View of Rainfall. *Physical Review*, 88, (1), 018701-1, 018701-4.
15. Satya N. Majumdar, Arnab Pal. (2014) Extreme value statistics of correlated random variables. CNRS, LPTMS, Orsay, Paris Sud & Raman Research Institute, India.
16. Castellanos, O., Manuel López, J., María Senties J., Anabitarte, E. (2013). Intermittency, avalanche statistics, and long-term correlations in a turbulent plasma. *The Journal of Statistical Mechanics: Theory and Experiment*. Departamento de Física Aplicada, Universidad de Cantabria & IFCA, Santander.
17. María Sancho, J. (1994). *Introducción a las ecuaciones de Langevin y Fokker-Planck*. (Universidad de Barcelona). Notas curso de doctorado, material no publicado.

9 Apéndice

```
% % SERIE ADITIVA.
%Generación de una serie temporal aleatoria simple (con ruido blanco)
x= linspace(1, 12, 100);
y=(2*x)+normrnd(0,1,1,100);
plot(x,y)
xlabel('Tiempo');
ylabel('Variable aleatoria');
title ('Serie temporal aditiva','FontName','Arial','FontSize', 13);
xlim([1,12])
figure
%% SERIE MULTIPLICATIVA.
z=(2*x).*normrnd(0,1,1,100);
% % %z=sin(2*pi*x/10+.5*pi).*normrnd(0,1,1,100);
plot(x,z,'red')
xlabel('Tiempo');
ylabel('Variable aleatoria');
title ('Serie temporal multiplicativa','FontName','Arial','FontSize', 13);
xlim([1,12])

%% SERIE ALEATORIA.
x= linspace(1, 100, 100);
y=sin(2*pi*x+.5*pi)+normrnd(0,1,1,100);
plot(x,y,'color',[0.000000 0.660000 0.470000])
xlabel('Tiempo');
ylabel('Variable aleatoria');
title ('Serie temporal estacionaria','FontName','Arial','FontSize', 13)
%xlim([1,12])
figure
% % SERIE NO ESTACIONARIA.
z=(x.^2).*normrnd(0,1,1,100);
plot(x,z,'color',[0.450000 0.310000 0.590000])
xlabel('Tiempo');
ylabel('Variable aleatoria')
title ('Serie temporal no estacionaria','FontName','Arial','FontSize', 13)

%% DISTRIBUCIONES GEV.
% Distribución generalizada de los valores extremos.
x = linspace(-4,8,1000);

y1 = gevpdf(x,0,1,0); %Gumbell
y2 = gevpdf(x,-.5,1,0); % Weibull
y3 = gevpdf(x,.5,1,0); % Frechet
%plot(x,y1,'-', x,y2,'-', x,y3,'-')
plot(x,y1,'color',[0 0 1]) %Azul
hold on
plot(x,y2,'color',[0 0.5 0]) % Verde
hold on
plot(x,y3,'color',[1.000000 0 0.000000]) %red
leyenda1='Gumbel';
leyenda2='Weibull';
leyenda3='Frechet';
leyendaf = legend(leyenda1,leyenda2,leyenda3);
set(leyendaf,'FontName','arial','FontUnits','points','FontSize',14,...
    'FontWeight','normal','FontAngle','normal');
set(gca,'fontsize',12)
```

```

xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones Generalizadas de Valores Extremos.','FontName','Arial','FontSize', 16);

figure

%Dibujo los logaritmos de las distribuciones y comparo con gaussiana l0= normpdf(X,mu,sigma);.
normal= normpdf(x,0,1);
x = linspace(-100,0,1000);
l0= log(normal);
ls = [];
ls(1,:) = log(y1);
ls(2,:) = log(y2);
ls(3,:) = log(y3);

iMaxs = [];
vMaxs = []
[vMaxL0, iMaxL0] = max(l0);
[vMaxs(1), iMaxs(1)] = max(ls(1,:));
[vMaxs(2), iMaxs(2)] = max(ls(2,:));
[vMaxs(3), iMaxs(3)] = max(ls(3,:));

for i = 1:3
    ls(i,:) = ls(i,:) * (vMaxL0 / vMaxs(i));
    dif = abs(iMaxs(i) - iMaxL0);
    [~, final] = size(l0);
    destino = zeros([1 final]);
    if (iMaxs(i) > iMaxL0)
        destino(1 : final - dif) = ls(i, 1 + dif : final);
    else
        destino(1 + dif : final) = ls(i, 1 : final - dif);
    end;
    ls(i, :) = destino(:);
end
% a=[0.9058    0.6557    0.0344];

%plot(x,ls(1,:),':');

%plot(x,l0,'--black',x,ls(1:3,:),'-blue', x,ls(2,:),'-red', x,ls(3,:),'-yellow')
plot(x,l0,'color',[0.550000 0.000000 0.550000]) % Gaussiana lila
hold on
plot(x,ls(1:3,:), 'color',[0 0 1])% Gumbel blue
hold on
plot(x,ls(2,:), 'color',[0 0.5 0]) %weibull green
hold on
plot(x,ls(3,:), 'color',[1 0 0]) %Frechet red

leyenda4='Gaussiana';
leyenda5='Gumbel';
leyenda6='Weibull';
leyenda7='Frechet';
leyenda = legend(leyenda4,leyenda5,leyenda6,leyenda7);
set(leyenda,'FontName','arial','FontUnits','points','FontSize',14,...
    'FontWeight','normal','FontAngle','normal');
xlabel('Valores');
ylabel('Logaritmo Densidad de Probabilidad');
title ('Logaritmo de las distribuciones GEV','FontName','Arial','FontSize', 16);
set(gca,'fontsize',12)

```

```
ylim([-2 -0.8])
xlim([-90 -40])
```

```
%%COMPARACION DISTRIBUCION GUMBEL CON DIFERENTES PARAMETROS.
```

```
%Primero represento la distribucion de Gumbel modificando el parámetro de
%escala (sigma) y de localización (mu).
```

```
x = linspace(-5,20,1000);
k=0; % Común a todas las Gumbel.
y1 = gevpdf(x,k,2,0.5);
y2 = gevpdf(x,k,2,1);
y3 = gevpdf(x,k,3,1.5);
y4= gevpdf(x,k,4,3);
plot(x,y1,'green', x,y2,'blue', x,y3,'black',x,y4,'red')%
legend( '\sigma=2.0 \mu=0.5' ,'\sigma=2.0 \mu=1.0', '\sigma=3.0 \mu=1.5', '\sigma=4.0 \mu= 3.0')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Gumbel','FontName','Arial','FontSize', 14);
```

```
% En segundo lugar, represento para mismo parametro de forma y de localización. Vario solo la esca
```

```
x = linspace(-2,10,1000);
mu=2;
y1 = gevpdf(x,k,0.5,mu);
y2 = gevpdf(x,k,1,mu);
y3 = gevpdf(x,k,1.5,mu);
y4= gevpdf(x,k,3,mu);
plot(x,y1,'green', x,y2,'blue', x,y3,'black',x,y4,'red')
%legend( '\sigma=0.5 ' ,'\sigma=1.0 ' ,'\sigma=1.5', '\sigma=3.0 ')
legend( '\sigma=0.5 \mu= 2.0' ,'\sigma=1.0 \mu=2.0', '\sigma=1.5 \mu=2.0', '\sigma=3.0 \mu=2.0')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Gumbel','FontName','Arial','FontSize', 14);
```

```
%%COMPARACION DISTRIBUCION FRECHET CON DIFERENTES PARAMETROS.
```

```
%Primero represento la distribucion de Frechet modificando el parámetro de
%forma k.
```

```
x = linspace(-10,10,100);
mu= 0;
sigma=10;

y1 = gevpdf(x,0.5,sigma,mu);
y2 = gevpdf(x,1,sigma,mu);
y3= gevpdf(x,1.5,sigma,mu);
y4 = gevpdf(x,2,sigma,mu);
y5=gevpdf(x,3,sigma,mu);
plot(x,y1,'green', x,y2,'blue', x,y3,'black',x,y4,'red',x,y5,'magenta')
legend( '\xi=0.5', '\xi=1.0', '\xi=1.5', '\xi=2.0', '\xi=3.0')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Fréchet','FontName','Arial','FontSize', 13);
```

```
%Represento la distribución de Fréchet variando el parámetro de escala.
```

```
x1 = linspace(-1.5,3,100);
mu= 0;

y1 = gevpdf(x1,0.5,1,mu);
y2 = gevpdf(x1,1,1,mu);
y3= gevpdf(x1,1.5,1,mu);
```

```

y4 = gevpdf(x1,2.0,2,mu);

plot(x1,y1,'red', x1,y2,'blue', x1,y3,'black',x1,y4,'green')
legend( '\xi=0.5 \sigma=1.0', '\xi=1.0 \sigma=1.0', '\xi=1.5 \sigma=1.0', '\xi=2.0 \sigma=1.0')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Fréchet','FontName','Arial','FontSize', 13);
figure

x2 = linspace(-2,2.5,100);

y5 = gevpdf(x2,0.5,2,mu);
y6 = gevpdf(x2,1,2,mu);
y7= gevpdf(x2,1.5,2,mu);
y8=gevpdf(x2,2.0,2,mu);
plot(x2,y5,'red', x2,y6,'blue', x2,y7,'black',x2,y8,'green')
legend( '\xi=0.5 \sigma=2.0', '\xi=1.0 \sigma=2.0', '\xi=1.5 \sigma=2.0', '\xi=2.0 \sigma=2.0')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Fréchet','FontName','Arial','FontSize', 13);

%%COMPARACION DISTRIBUCION WEIBULL CON DIFERENTES PARAMETROS FORMAS.
clc,clear all, close all
% Primero represento la distribución de Weibull variando solamente el
% parámetro de forma k.
x = linspace(-2,2,500);
mu= 0;
sigma=1;

y1 = gevpdf(x,-0.5,sigma,mu);
y2 = gevpdf(x,-0.9,sigma,mu);
%y3= gevpdf(x,-1.5,sigma,mu);
%y4 = gevpdf(x,-1.75,sigma,mu);
%y5=gevpdf(x,-3,sigma,mu);
plot(x,y1,'green', x,y2,'blue', x,y3,'black',x,y4,'red')%x,y5,'magenta')
legend( '\xi =0.5', '\xi =1.0', '\xi =1.5', '\xi =2.0', '\xi=3.0')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribución de Weibull.','FontName','Arial','FontSize', 14);

%plot(x,y1,'blue', x,y2,'green', x,y3,'black',x,y4,'red',x,y5,'yellow')

%%Represento la distribución de Weibull variando el parámetro de escala.
x1 = linspace(-6,6,1000);
mu= 0;
sigma1=1;
y1 = gevpdf(x1,-0.3,sigma1,mu);
y2 = gevpdf(x1,-0.5,sigma1,mu);
y3= gevpdf(x1,-0.8,sigma1,mu);
y4 = gevpdf(x1,-1.0,sigma1,mu);

plot(x1,y1,'color',[0.860000 0.080000 0.240000])
hold on
plot(x1,y2,'color',[1.000000 0.840000 0.000000])
hold on
plot(x1,y3,'color',[0 0 1])
hold on

```

```

plot(x1,y4,'color',[0 0.5000000 0.000000])
legend( '\xi =-0.3 \sigma=1.0','\xi =-0.5 \sigma=1.0','\xi =-0.8 \sigma=1.0','\xi=-1.0 \sigma=1.0'
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Weibull','FontName','Arial','FontSize', 13);
xlim([-6 6])
figure

x2 = linspace(-6,6,100);
sigma2=2.0;
y5 = gevpdf(x2,-0.3,sigma2,mu);
y6 = gevpdf(x2,-0.5,sigma2,mu);
y7= gevpdf(x2,-0.8,sigma2,mu);
y8=gevpdf(x2,-1.0,sigma2,mu);
%plot(x2,y5,'red', x2,y6,'blue', x2,y7,'black',x2,y8,'green')
plot(x2,y5,'color',[0.860000 0.080000 0.240000])
hold on
plot(x2,y6,'color',[1.000000 0.840000 0.000000])
hold on
plot(x2,y7,'color',[0 0 1])
hold on
plot(x2,y8,'color',[0 0.5000000 0.000000])

legend( '\xi =-0.3 \sigma=2.0','\xi =-0.5 \sigma=2.0','\xi =-0.8 \sigma=2.0','\xi =-1.0 \sigma=2.0'
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Weibull','FontName','Arial','FontSize', 13);
% Represento la variabilidad con el parámetro de forma. Fijo sigma y mu.
x3 = linspace(-3,3,1000);
mu= 0;
sigma=1;
j1 = gevpdf(x3,-0.3,sigma,mu);
j2 = gevpdf(x3,-0.9,sigma,mu);
j3= gevpdf(x3,-1.5,sigma,mu);
j4 = gevpdf(x3,-1.1,sigma,mu);

plot(x3,j1,'red', x3,j2,'blue', x3,j3,'black',x3,j4,'green')
legend( '\xi =-0.3 ','\xi =-0.5','\xi =-0.8 ','\xi =-1.0 ')
xlabel('Valores');
ylabel('Densidad de Probabilidad');
title ('Distribuciones de Weibull','FontName','Arial','FontSize', 13);

%% DISTRIBUCIÓN GENERALIZADA DE PARETO.
%Forma general de la distribución generalizada de Pareto (GPD)
x = linspace(0,10,1000);

y1 = gppdf(x,-0.35,1,0);
y2 = gppdf(x,0,1,0);
y3 = gppdf(x,0.35,1,0);

plot(x,y1,'color',[0 0 1])
hold on
plot(x,y2,'color',[0 0.5 0 ])
hold on
plot(x,y3,'color',[1 0 0])
legend( '\xi=-0.35','\xi=0','\xi=0.35')

```

35

36

```

ajusteNorm = ajuste*1/an;
bar(histMout(1:escala),ajusteNorm);
plot(histMout(1:escala), ajusteNorm, 'r');

%%CONVERGENCIA PARAMETRO FORMA FRECHET.
ji=[ 1.0022 0.9911 0.9996 0.998 1.005 1.002 0.9998 0.9999 1.002 1.00 0.9997 1.002 1.002];
N=[ 5 10 50 100 500 1e3 2e3 5e3 7e3 8e3 1e4 2e4 3e4 ];
plot(N,ji,'-.*b')
xlabel('Block Size','FontSize', 13)
ylabel('Parámetro de forma','FontSize', 13)

%% BLOCK MAXIMA WEIBULL.
clc,clear all, close all
tic
NNminimo = 1e5; % número de bloques
Maximos_size = 0;

NN = 1e4;
n0 = 1;
N = 100; % Tamaño de bloque

Maximos=zeros(n0*NNminimo,1 );
while NNminimo > Maximos_size
    MM = unifrnd(0,1,[N,NN] );
    v = max(MM); % vector columna de máximos
    Maximos( (Maximos_size+1):(Maximos_size+numel(v)) ) = v;
    Maximos_size = Maximos_size + numel(v);

end
toc
zoom=1;
num_bins=1e7;
escala=num_bins/zoom;
[histFreq,histMout]=hist((Maximos),num_bins);

xlabel('x');
ylabel('PDF');
hold on
% Hago el ajuste:
an=N;
bn=1;
PD= fitdist(Maximos,'gev')
valGumbel = pdf(PD,histMout);
valGumbelNorm=(1/an).*((valGumbel-bn));
bar(histMout,valGumbelNorm);
plot(histMout,valGumbelNorm,'r');

% histogram(Maximos,'Normalization','pdf')
% hold on
% PD=fitdist(Maximos, 'gev');
% plot(0:0.1:1, pdf(PD,0:.1:1))
% %
xlabel('x');
ylabel('PDF');
ylim([0 1])

%%CONVERGENCIA PARAMETRO FORMA WEIBULL.

```



```

ji2=[ -1.001 -0.9985 -1.002 -0.9935 -0.9971    -0.9849 -1.001 -1.0011   -0.9592    -0.9949   -0.8952
N2=[ 5 10 50  100 500  1e3 2e3  5e3 7e3 8e3 1e4 2e4 3e4 ];
plot(N2,ji2,'-.*g')
xlabel('Block Size','FontSize', 13)
ylabel('Parámetro de forma','FontSize', 13)

%% EFECTO CORRELACION.
% Fichero con datos interior cilindro (independientes)
fileID1=fopen('flujoabs_1054.txt','r')
formatSpec = '%f';
data1=fscanf(fileID1,formatSpec)
fclose(fileID1)
%Fichero con datos región bordes (correlacionados)
fileID=fopen('flujoabs_1067.txt','r')
formatSpec = '%f';
data2=fscanf(fileID,formatSpec)
fclose(fileID)

figure;
T=1:1:2e5;
subplot(2,1,1)
plot(T,data1,'k-')
xlabel('Tiempo')
ylabel('ABS(Flujo)')
title('Serie con datos independientes','FontName','Arial','FontSize', 16);
set(gca,'fontsize',12)
xlim([0 1e4])

subplot(2,1,2)
plot(T,data2,'k-')
xlabel('Tiempo')
ylabel('ABS(Flujo)')
title('Serie con correlación temporal','FontName','Arial','FontSize', 16);
set(gca,'fontsize',12)
xlim([0 1e4])

%% DIAGRAMA DE BIFURCACIONES DEL MAPA LOGISTICO.
% Iteramos la ecuación logística:  $x = \mu * x * (1 - x)$ 
clc,clear all, close all
hold on
for mu=1:0.0025:4
    x=0.1; % CI
    MU=mu*ones(1,200);
    for i=1:100
        x=mu*x*(1-x);
    end
    X=zeros(1,0);
    for j=1:200
        x=mu*x*(1-x);
        X=[X,x];
    end
    plot(MU,X,'.black','Markersize',2)
end
ylabel('x_n','FontSize',12);
xlabel('Parámetro de Control: \mu','FontSize',12)
title('Diagrama de Bifurcación','FontName','Arial','FontSize', 16)
set(gca,'fontsize',13)

```

```

%% EXPONENTE DE LYAPUNOV VS PARAMETRO DE CONTROL.
% Calculo los exponentes de Lyapunov para la ecuación logística.
% Uso el método derivativo.
close all
lyapunov=zeros(1,1000);
i=0;
for mu=3:0.001:4
    xn1=0.1; %CI
    l=0;
    i=i+1;
    for j=1:10000
        xn=xn1;

        xn1=mu*xn*(1-xn); %ecuacion logistica

        if j>300
            l=l+log(abs(mu-2*mu*xn1));%aplico metodo derivativo.
        end
    end
    l=l/10000; %Divido entre el numero de iteracciones.
    lyapunov(i)=l;
end
mu=3:0.001:4;
plot(mu,lyapunov);
hold on
plot([3,4],[0,0]);
xlabel('\mu');
ylabel('\lambda');
title('Exponente de Lyapunov vs Parámetro de control','FontName','Arial','FontSize', 16);
set(gca,'fontsize',13)

%% RUIDO GAMMA ESTOCÁSTICO GAUSSIANO.
dt = 1e-1;
tfinal = 1e5;
tau=0.1; % tiempo de correlación
rng(1); % inicializo la semilla aleatoria.
numts = tfinal * (1/dt);
t2i = @(t) int32(t * (1/dt) + 1);

gamma= zeros(1, numts);

Maximos = zeros(1,1);
for i=1
    gamma(t2i(0))= 1;

    for t=0:dt:tfinal
        gamma(t2i(t+dt)) = gamma(t2i(t))*(1-(dt/tau))+(randn/tau);
    end
end
pd = fitdist(gamma,'Normal');
histogram(gamma,600,'Normalization','pdf')
hold on
plot(-40:0.01:40, pdf(pd, -40:0.01:40))
set(gca,'fontsize',12)
xlim([-40 40])
title('Ruido Gaussiano de Ornstein-Uhlenbeck','FontName','Arial','FontSize', 16);
xlabel('\Gamma')
ylabel('PDF')

```

```

%% BLOCK MAXIMA PROCESO ORNSTEIN-UHLENBECK.
%Genero series estocasticas con ruido correlacionado.
tau=0.1;
dt = 1e-1;
t = 0:dt:1e2; %vector de tiempos
nbloques=1e4;
x = zeros(1,length(t));
g =zeros(1,length(t));
maximo=zeros(1,nbloques);
rng(1);
for k = 1:nbloques

for i = 1:length(t)-1
    g(i+1) = g(i)*(1-(dt/tau))+(randn/tau);
    x(i+1)=x(i)+g(i)*dt+0.5*((g(i))^2)* ((dt)^2);
    maximo(k)=max(x);

end
end
gevfit(maximo) %estimación parámetros

%% CONVERGENCIAS DEL PARAMETRO DE FORMA PARA RUIDO CORRELACIONADO OU.
%TC=10
ji1=[5.0609 3.1411 3.5012 2.4849 1.3180 0.9705 0.7701 0.3897 0.1968 0.1467 0.1297 0.0676 0.0676];
N1=[5 10 50 60 80 90 100 200 500 800 1e3 2e3 3e3 5e3 1e4 ];
plot(N1,ji1,'-.*g')

hold on
%TC=100
ji2=[ 5.1381 5.0342 4.8628 4.7331 4.6118 3.9925 3.3696 2.8805 1.4265 0.9898 0.5129 0.3546 0.2001];
N2=[ 5 10 50 100 200 500 600 700 800 1e3 2e3 3e3 5e3 7e3 1e4 2e4 3e4 5e4 1e5 ];%5e3 7e3 8e3 1e5
plot(N2,ji2,'-.*r')
hold on
%TC=1000
ji4=[ 3.3960 4.8927 3.7582 4.5400 4.0863 4.8162 4.3923 3.5127 3.7981 3.5396 2.1730 1.8978 0.9978];
N4=[ 5 10 50 100 200 500 800 1e3 2e3 5e3 7e3 1e4 3e4 5e4 7e4 1e5 2e5 5e5 7e5 1e6];
plot(N4,ji4,'-.*b')

legend('\tau_{c}=10 ', '\tau_{c}=100', '\tau_{c}=1000')
title('Convergencia \xi para el proceso OU correlacionado ', 'FontName','Arial','FontSize', 16);
xlabel('Tamaño de bloque N','FontSize', 13)
ylabel('Parámetro de forma \xi','FontSize', 13)
set(gca,'fontsize',13)

%%CONVERGENCIA PARA TC PEQUEÑOS:
%TC=1
ji=[ 4.7242 1.2947 0.8549 0.7094 0.5874 0.5975 0.2613 0.1773 0.1315 0.1255 0.1021 0.0506 0.0366 0.0366];
N=[ 5 6 7 8 9 10 20 25 28 30 35 45 50 55 57 59 60 62 65];
plot(N,ji,'-.*b')
hold on
legend('\tau_{c}=1')
title('Convergencia de \xi para \tau_{c} pequeños ', 'FontName','Arial','FontSize', 16);
xlabel('Tamaño de bloque N','FontSize', 13)
ylabel('Parámetro de forma \xi','FontSize', 13)
set(gca,'fontsize',12)

```